ARI Working Papers

Selection and Classification Technical Area

1983-1990

These working papers are published in order to archive material that was not included in other ARI publications. The material contained herein may not meet ARI's usual scientific or professional standards for publication.

July 2001

United States Army Research Institute for the Behavioral and Social Sciences

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) July 2001	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) 1983-1990
4. TITLE AND SUBTITLE ARI Working Papers: Selection and Classification Technical Area, 1983-1990		5a. CONTRACT OR GRANT NUMBER
		5b. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Arabian, J.M., Wise, L.L., Young, W., Busciglio, H.H., Campbell, J.P., Chia, W.J., Czarnolewski, M.Y., Martin, C.J., Eaton, N.K., Wing, H., Mitchell, K., Edwards, D.C., Claudy, J.J., Fischl, M.A., Rumsey, M.G., Gast, I.F., Grafton, F.C., Smith, E.P., Olson, D.M., Hanser, L.M., Walker, C.B., Harris, J.H., Osborne, W., Park, R.K., Rossmeissl, P.G., Silva, J.M., Birnbaum, A.E., Austin, K.C., Knapp, D.J., McHenry, J.J., Owens-Kurtz, C., Peterson, N., and Szenas, P.L.		5c. PROJECT NUMBER
		5d. TASK NUMBER
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-PO 5001 Eisenhower Avenue Alexandria, VA 22333-5600		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER WP SCTA

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

ARI working papers were originally unofficial documents intended for limited distribution to obtain comments. These working papers are being archived in order to preserve material that was not included in other ARI publications. The material contained herein may not meet ARI's usual scientific or professional standards for publication.

14. ABSTRACT *(Maximum 200 words)*:

Forty-one working papers dealing with selection and classification, aptitude areas, the SQT, the AFQT, job performance, testing, Project A, EEO, synthetic validation, officer selection, the ASVAB, TOW gunnery, spatial tests, CAT-ASVAB, and psychomotor tests.

15. SUBJECT TERMS

Selection, aptitude areas, SQT, AFQT, job performance, testing, Project A, EEO, synthetic validation, officer selection, ASVAB, TOW gunnery, spatial tests, CAT-ASVAB, psychomotor tests, Officer Selection Battery, armor, concurrent validation, classification

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	1729	David W. Witter (703) 617-0324

NAME________________________________

TASK MATCHING EXERCISE

Task List for 11B
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Prepare a dragon for firing	____	____	____
2.	Techniques of movement in urban terrain	____	____	____
3.	Call for/adjust indirect fire	____	____	____
4.	Select fire team/scout squad overwatch position	____	____	____
5.	Set headspace and timing on a caliber .50 machinegun	____	____	____
6.	Place an AN/PVS-5 (night vision goggles) into operation	____	____	____
7.	Zero AN/PVS-4 to an M16A1 rifle	____	____	____
8.	Select hasty firing positions in urban terrain	____	____	____
9.	Operate radio set AN/PRC-77 or AN/PRC-25 (AN/GRC-160 or AN/GRC-125 assembled for manpack operations)	____	____	____
10.	Collect/report information --SALUTE	____	____	____
11.	Recognize and identify friendly and threat armored vehicles	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
13.	Identify terrain features on a map	____	____	____
14.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
15.	Load, reduce a stoppage, and clear an M60 machinegun	____	____	____
16.	Prepare a range card for an M60 machinegun	____	____	____
17.	Engage targets with an M72A2 LAW	____	____	____
18.	Engage enemy targets with hand grenades	____	____	____
19.	Install and fire/recover an M18A1 Claymore mine	____	____	____
20.	Move under direct fire	____	____	____
21.	Move over, through, or around obstacles (except minefields)	____	____	____
22.	Camouflage yourself and your individual equipment	____	____	____
23.	Put on, wear, and remove your M17-series protective mask with hood	____	____	____
24.	Put on and wear protective clothing	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
25.	Perform cardiopulmonary resuscitation (CPR) on an adult using the one-man method	____	____	____
26.	Put on a field or pressure dressing	____	____	____
27.	Navigate from one point on the ground to another point	____	____	____
28.	Establish an observation post	____	____	____
29.	Conduct day and night surveillance without the aid of electronic devices	____	____	____
30.	Administer nerve agent antidote to self (self-aid)	____	____	____

NAME________________________________

TASK MATCHING EXERCISE

Task List for 63B
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Adjust clutch pedal free travel	____	____	____
2.	Maintain assigned toolkit	____	____	____
3.	Repair electrical wiring (truck, cargo, 1-1/4 ton, 4x4)	____	____	____
4.	Replace air hydraulic cylinder	____	____	____
5.	Replace radiator (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
6.	Replace service brakes (truck, utility, 1/4 ton, 4x4)	____	____	____
7.	Replace fuel pump (truck, cargo, 2-1/2, ton, 6x6)	____	____	____
8.	Replace wheel bearings (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
9.	Tow disabled vehicle with 5 ton wrecker	____	____	____
10.	Perform annual PMCS (inspect and road test) (truck, cargo, 5 ton, 6x6)	____	____	____
11.	Troubleshoot brake system malfunctions (truck, cargo, 1-1/4 ton, 4x4)	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
12.	Troubleshoot electrical system (truck, cargo, 5 ton, 6x6)	____	____	____
13.	Troubleshoot fuel system malfunctions (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
14.	Troubleshoot service brake malfunctions (truck, utility, 1/4 ton, 4x4)	____	____	____
15.	Replace starter	____	____	____
16.	Administer first aid to a nerve agent casualty (buddy-aid)	____	____	____
17.	Camouflage equipment	____	____	____
18.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
19.	Determine grid coordinates	____	____	____
20.	Determine magnetic azimuth using compass	____	____	____
21.	Load, reduce stoppage, clear M16A1 rifle	____	____	____
22.	Use challenge and password	____	____	____
23.	Put on and wear M17 series protective mask with hood	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
24.	Put on and wear protective clothing in accordance with established (MOPP) levels	____	____	____
25.	Put on a field or pressure dressing	____	____	____
26.	Perform expedient repairs	____	____	____
27.	Troubleshoot engine cooling system	____	____	____
28.	Troubleshoot engines	____	____	____
29.	Troubleshoot steering system	____	____	____
30.	Slave start disabled vehicle	____	____	____

NAME________________________________

TASK MATCHING EXERCISE

Task List for 71L
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Type a basic comment to a disposition form	____	____	____
2.	Type a military letter	____	____	____
3.	Type a joint messageform (DD Form 173/1)	____	____	____
4.	Prepare a requisition for publications and/or blank forms--AUTODIN (DA Form 4569)	____	____	____
5.	Type military orders	____	____	____
6.	Type a memorandum	____	____	____
7.	Dispatch outgoing distribution	____	____	____
8.	Establish functional files	____	____	____
9.	File documents/correspondence	____	____	____
10.	Type a second or subsequent comment to a disposition form	____	____	____
11.	Type straight copy material	____	____	____
12.	Assemble correspondence	____	____	____
13.	Receipt/transfer classified material	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
14.	Safeguard "FOR OFFICIAL USE ONLY" (FOUO) material	____	____	____
15.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
16.	Load, reduce a stoppage, and clear an M16A1 rifle	____	____	____
17.	Camouflage yourself and your individual equipment	____	____	____
18.	Practice noise, light, and litter discipline	____	____	____
19.	Determine the grid coordinates of a point on a military map using the military grid reference system	____	____	____
20.	Determine a magnetic azimuth using a compass	____	____	____
21.	Put on a field or pressure dressing	____	____	____
22.	Maintain an M17 series protective mask	____	____	____
23.	Put on and wear an M17 series protective mask with hood	____	____	____
24.	Put on and wear protective clothing in accordance with established mission-oriented protective posture (MOPP) levels	____	____	____
25.	Know your rights and obligations as a prisoner of war	____	____	____
26.	Administer nerve agent antidote to self (self-aid)	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
27.	Receive, maintain, and control office equipment	______	______	______
28.	Control expendable/non-expendable supplies	______	______	______

APPENDIX C

JOB ACTIVITY INSTRUMENT

Name: ______________________

Rank: ______________________

MOS: ______________________

MOS JOB ACTIVITY QUESTIONNAIRE

Listed below are 53 job activities. For each activity, we would like you to make the following four ratings:

1. Is the activity a **PART OF THE JOB** for E3 and E4 soldiers in the MOS you are rating?

   _N_ = No (skip ratings 2 - 4 and go to the next activity)

   _Y_ = Yes (continue with ratings 2 - 4)

2. How **IMPORTANT** or critical is the activity for successful performance for E3 and E4 soldiers in the MOS?

   _5_ = Extremely high importance for E3 and E4 soldiers. The activity is very central to the job. (Only the most critical activities should get this rating.)

   _4_ = High importance. The activity is central to the job, but it is not one of the most critical activities.

   _3_ = Moderate importance. The activity is part of the job, but it is not central to the job.

   _2_ = Low importance. The activity is part of the job, but there are many activities that are more important.

   _1_ = Very minor importance. The activity is part of the job, but it is not at all important for successful job performance.

3. How **FREQUENTLY** is the activity performed by E3 and E4 soldiers in the MOS?

   _3_ = Very frequently. The activity is performed much more frequently than most other activities.

   _2_ = Frequently. The activity is performed about as often as other activities.

   _1_ = Occasionally or infrequently. The activity is performed much less frequently than most other activities.

4. How **DIFFICULT** and demanding is it for E3 and E4 soldiers in the MOS to perform this activity well?

- 3 = Very difficult. It takes considerable skill and motivation to perform this activity well. Few soldiers can perform this activity.
- 2 = Moderately difficult. It requires some skill and motivation to perform this activity well, but most soldiers are eventually able to learn the activity.
- 1 = Routine. The activity is relatively easy and requires little skill. Almost all soldiers learn the activity very quickly and are able to perform the activity with little difficulty.

	Part of Job?	Importance	Frequency	Difficulty

Leadership/Teamwork

	Part of Job?	Importance	Frequency	Difficulty
1. Work in a team -- participate in work or combat teams by helping resolve conflicts, keep the team moving toward its goals, and maintain morale of other group members.	____	____	____	____
2. Lead a team -- lead or direct a team or unit as it tries to accomplish a prescribed mission or goal.	____	____	____	____
3. Support/counsel peers -- support and/or counsel peers when they have difficulty.	____	____	____	____
4. Support/counsel subordinates -- counsel subordinates when they have difficulty.	____	____	____	____
5. Coach peers -- instruct/train peers on specific job tasks.	____	____	____	____
6. Coach subordinates -- instruct/train subordinates on specific job tasks.	____	____	____	____

Communication

	Part of Job?	Importance	Frequency	Difficulty
7. Make oral reports (to individuals) -- make oral reports to other individuals.	____	____	____	____
8. Make oral reports (to groups) -- give oral briefings to a group.	____	____	____	____
9. Relay oral instructions -- relay oral instructions on job tasks or job procedures to other individuals or to groups.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
10 <u>Ask questions</u> -- ask frequent questions, as in interviewing and investigating.	____	____	____	____
11. <u>Record information</u> -- record messages via handwriting or using a keyboard.	____	____	____	____
12. <u>Write brief messages</u> -- write brief messages (e.g., memos, day-to-day instructions, business letters, etc.)	____	____	____	____
13. <u>Write longer reports</u> -- compose and write longer formal reports or articles (i.e., longer than one or two pages).	____	____	____	____
Use Information				
14. <u>Monitor/interpret verbal messages</u> -- monitor, identify, and/or interpret verbal messages (oral or written, obtained from radio, teletype, computer terminal, correspondences, etc.)	____	____	____	____
15. <u>Recall verbal information</u> -- recall verbal information. (That is, to perform the required job tasks, the individual must rely on memory for necessary verbal information that was communicated more than a few hours ago. However, this item does <u>not</u> refer to information presented during formal training.)	____	____	____	____
16. <u>Monitor/interpret numerical information</u> -- monitor, identify, and/or interpret numerical/quantitative information that is presented via written reports, radio, CRT, or other electronic equipment.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
17. <u>Recall numerical information</u> -- recall numerical/quantitative information. (Again, "recall" refers to the fact that the job requires the use of memory to have the necessary information to perform the job tasks. However, it is <u>not</u> memory for training content but memory for information acquired as part of the job.)	____	____	____	____
18. <u>Monitor/interpret figural information</u> -- monitor, identify, and/or interpret figural information (e.g., CRT images, pictures, graphs, terrain features).	____	____	____	____
19. <u>Recall figural information</u> -- recall figural information (e.g., CRT images, picutres, graphs, terrain features).	____	____	____	____
20. <u>Follow oral directions</u> -- follow oral directions (e.g., for how to complete documents, move from point A to point B, or schedule repairs on a variety of equipment).	____	____	____	____
21. <u>Follow written directions</u> -- follow written directions/instructions as part of normal job duties (e.g., follow written orders, SOP protocols, or technical manuals.)	____	____	____	____
Perceptual Judgements				
22. <u>Judge size and distance</u> -- make judgements of relative size and distance, as in which of two objects is closer/larger or the distance to a target.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
23. <u>Judge location</u> -- orient oneself relative to location and direction from physical objects or terrain features and maintain the proper orientation while moving from point to point.	____	____	____	____
24. <u>Judge paths of moving objects</u> -- make judgements about the relative position of moving objects (e.g., Where will two vehicles converge? As the truck cab moves, where will the trailer go?).	____	____	____	____
Problem Solving/Troubleshooting				
25. <u>Solve electrical system problems</u> -- troubleshoot electrical equipment problems (i.e., identify the problem and suggest solution).	____	____	____	____
26. <u>Solve mechanical system problems</u> -- troubleshoot mechanical equipment problems (i.e., identify the problem and suggest solution).	____	____	____	____
27. <u>Solve logistical problems</u> -- identify and solve logistical problems.	____	____	____	____
28. <u>Solve tactical manuever problems</u> -- identify and solve tactical maneauver problems (e.g., positioning of teams in the field, countering moves of enemy).	____	____	____	____
29. <u>Solve administrative problems</u> -- identify and solve administrative problems (e.g., correct record keeping errors, personnel scheduling).	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
30. <u>Solve leadership problems</u> -- identify and solve leadership/supervision problems (e.g., involving discipline, training needs, conflicts between people, motivation of team members).	____	____	____	____
31. <u>Solve medical problems</u> -- identify and solve physical health problems (i.e., problems that require special training beyond basic first aid).	____	____	____	____
32. <u>Solve communication problems</u> -- identify and solve communication problems (i.e., not technical/equipment problems, but problems resulting from inaccurate or non-extistent commuinication).	____	____	____	____
Operate Equipment				
33. <u>Operate precision hand-held equipment</u> -- operate hand-held equipment requiring great precision (e.g., syringe, calipers, soldering or welding equipment).	____	____	____	____
34. <u>Operate hand-held tools</u> -- operate hand-held equipment that does not require great precision (e.g., hammer, wrench).	____	____	____	____
35. <u>Operate hand-held power equipment</u> -- operate hand-held power assisted equipment (e.g., electric saw, electric wrench).	____	____	____	____
36. <u>Operate large power equipment</u> -- operate large power assisted equipment (e.g., forklift, bulldozer, backhoe).	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
37. Operate full keyboard -- operate full typewriter or computer keyboard.	____	____	____	____
38. Operate numeric keyboard -- operate a numeric keyboard (i.e., just 10 basic keys, 0-9).	____	____	____	____
Adjust and Control				
39. Adjust control device using one limb -- make adjustments in control devices by using just one hand or one foot (e.g., a dial, a lever, or a foot pedal).	____	____	____	____
40. Adjust control device using multiple limbs -- make adjustments in control devices by using more than one limb (e.g., twisting a cylinder with one hand while setting a dial with the other, or depressing a footpedal while moving equipment into proper alignment with hands).	____	____	____	____
Drive				
41. Drive tracked vehicle -- drive heavy vehicles (tracked).	____	____	____	____
42. Drive heavy wheeled vehicle -- drive heavy vehicles (wheeled).	____	____	____	____
43. Drive light wheeled vehicle -- drive light wheeled vehicles.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
Aiming				
44 Aim: stationary target -- aim at a stationary object.	____	____	____	____
45. Aim: moving target -- manually track a moving target.	____	____	____	____
Physical Actions				
46. Walk long distances -- walk long distances carrying significant weight.	____	____	____	____
47. Run short distances -- run short to middle distances.	____	____	____	____
48. Push, pull, lift heavy weights -- push, pull, or lift heavy weights.	____	____	____	____
49. Throw objects -- throw hand-held objects.	____	____	____	____
50. Sort, fold, feed by hand -- use hands to fold objects, sort objects, or feed objects into a machine.	____	____	____	____
51. Make coordinated movements -- use well coordinated hand, arm, and upper body movements (e.g., as in setting up equipment, or moving hazardous material quickly).	____	____	____	____
52. Work long hours -- work long hours without rest.	____	____	____	____
53. Work under adverse conditions -- work under adverse or danagerous conditions.	____	____	____	____

Name ____

JOB ACTIVITY RATING EXERCISE
EVALUATION SHEET

Job Activity Definitions

1. How clear were the job activity definitions?

1	2	3	4	5	6	7
Not at all clear			Neither clear nor unclear			Very clear

2. How well did you understand the terms used in the job activity definitions?

1	2	3	4	5	6	7
There were many terms I did not understand			There were some terms I did not understand			There were no terms I did not understand

3. Which job activity was most difficult to judge? ____________________

4. Which job activity was least difficult to judge? ____________________

5. What percentage of your job was covered by these activities?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

If you answered less than 100%, what activities should be added?

____

____

____

____

____

____

Job Activity Ratings

1. How clear were the instructions for making the ratings?

1	2	3	4	5	6	7
Not at all clear			Neither clear nor unclear			Very clear

2. Were the ratings easy or difficult to make?

1	2	3	4	5	6	7
Very difficult			Neither easy nor difficult			Very easy

3. How confident did you feel about the accuracy of your ratings?

1	2	3	4	5	6	7
Not at all confident			Somewhat confident			Very confident

General Evaluation

1. Which of the following would you change (check all that apply):

_____ Job Activity Definitions

_____ Job Activity Rating Instructions and Procedures

2. What changes would you make?

____

____

____

____

____

____

____

____

MATCHING EXERCISE INSTRUCTIONS: MOS TASKS AND JOB ACTIVITIES

In this portion of the workshop, we are interested in finding out how well our list of 53 job activities can be matched up with tasks from the Soldier's Manual (SM) and from the Army Occupational Survey Program (AOSP) task list.

We have reviewed the SM and the AOSP and have selected 30 tasks that cover the range of responsibilities for Skill Level 1 and Skill Level 2 soldiers. Following each task, there are three blank lines. In those blanks, we want you to list the numbers of up to 3 job activities that best describe that task. You can find the numbers of the job activities on the "MOS Job Activity Questionnaire" that you completed during the previous exercise. This form provides the number, the name, and a brief description of each of the 53 job activities. Or, if you would prefer, the job activity names and numbers also can be obtained from the attached handout, "List of Job Activities" (which does not include any descriptions of the activities).

As you review the job activities, first find the job activity that best describes the task. Put the number of this job activity in the first column, under "1st Category/Activity." Then find the number of the job activity that provides the second best description of the task and mark this number under "2nd Category/Activity." Finally, do the same for the third job activity.

You do not have to fill in all three blanks for all of the tasks. You may find that only one or two job activities describe some tasks. Feel free to leave the remaining column(s) blank. If you find that none of the job activities describes a task, simply write "NONE" across the three blanks.

Also, you are not limited in the number of times that you can use one of the job activities. For example, if the same job activity describes all of the tasks on the Task List, you can use that activity for every task.

List of Job Activities

1. Work in a team
2. Lead a team
3. Support/counsel peers
4. Support/counsel subordinates
5. Coach peers
6. Coach subordinates
7. Make oral reports (to individuals)
8. Make oral reports (to groups)
9. Relay oral instructions
10. Ask questions
11. Record information
12. Write brief messages
13. Write longer reports
14. Monitor/interpret verbal messages
15. Recall verbal information
16. Monitor/interpret numerical information
17. Recall numerical information
18. Monitor/interpret figural information
19. Recall figural information
20. Follow oral directions
21. Follow written directions
22. Judge size and distance
23. Judge location
24. Judge paths of moving objects
25. Solve electrical system problems
26. Solve mechanical system problems
27. Solve logistical problems
28. Solve tactical maneuver problems
29. Solve administrative problems
30. Solve leadership problems
31. Solve medical problems
32. Solve communication problems
33. Operate precision hand-held equipment
34. Operate hand-held tools
35. Operate hand-held power equipment
36. Operate larger power equipment
37. Operate full keyboard
38. Operate numeric keyboard
39. Adjust control device using one limb
40. Adjust control device using multiple limbs
41. Drive tracked vehicle
42. Drive heavy wheeled vehicle
43. Drive light wheeled vehicle
44. Aim: stationary target
45. Aim: moving target
46. Walk long distances
47. Run short distances
48. Push, pull, lift heavy weights
49. Throw objects
50. Sort, fold, feed by hand
51. Make coordinated movements
52. Work long hours
53. Work under adverse conditions

NAME________________________________

TASK MATCHING EXERCISE

Task List for 11B
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Prepare a dragon for firing	____	____	____
2.	Techniques of movement in urban terrain	____	____	____
3.	Call for/adjust indirect fire	____	____	____
4.	Select fire team/scout squad overwatch position	____	____	____
5.	Set headspace and timing on a caliber .50 machinegun	____	____	____
6.	Place an AN/PVS-5 (night vision goggles) into operation	____	____	____
7.	Zero AN/PVS-4 to an M16A1 rifle	____	____	____
8.	Select hasty firing positions in urban terrain	____	____	____
9.	Operate radio set AN/PRC-77 or AN/PRC-25 (AN/GRC-160 or AN/GRC-125 assembled for manpack operations)	____	____	____
10.	Collect/report information --SALUTE	____	____	____
11.	Recognize and identify friendly and threat armored vehicles	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
13.	Identify terrain features on a map	____	____	____
14.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
15.	Load, reduce a stoppage, and clear an M60 machinegun	____	____	____
16.	Prepare a range card for an M60 machinegun	____	____	____
17.	Engage targets with an M72A2 LAW	____	____	____
18.	Engage enemy targets with hand grenades	____	____	____
19.	Install and fire/recover an M18A1 Claymore mine	____	____	____
20.	Move under direct fire	____	____	____
21.	Move over, through, or around obstacles (except minefields)	____	____	____
22.	Camouflage yourself and your individual equipment	____	____	____
23.	Put on, wear, and remove your M17-series protective mask with hood	____	____	____
24.	Put on and wear protective clothing	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
25.	Perform cardiopulmonary resuscitation (CPR) on an adult using the one-man method	____	____	____
26.	Put on a field or pressure dressing	____	____	____
27.	Navigate from one point on the ground to another point	____	____	____
28.	Establish an observation post	____	____	____
29.	Conduct day and night surveillance without the aid of electronic devices	____	____	____
30.	Administer nerve agent antidote to self (self-aid)	____	____	____

NAME______________________________

TASK MATCHING EXERCISE

Task List for 63B
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Adjust clutch pedal free travel	____	____	____
2.	Maintain assigned toolkit	____	____	____
3.	Repair electrical wiring (truck, cargo, 1-1/4 ton, 4x4)	____	____	____
4.	Replace air hydraulic cylinder	____	____	____
5.	Replace radiator (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
6.	Replace service brakes (truck, utility, 1/4 ton, 4x4)	____	____	____
7.	Replace fuel pump (truck, cargo, 2-1/2, ton, 6x6)	____	____	____
8.	Replace wheel bearings (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
9.	Tow disabled vehicle with 5 ton wrecker	____	____	____
10.	Perform annual PMCS (inspect and road test) (truck, cargo, 5 ton, 6x6)	____	____	____
11.	Troubleshoot brake system malfunctions (truck, cargo, 1-1/4 ton, 4x4)	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
12.	Troubleshoot electrical system (truck, cargo, 5 ton, 6x6)	____	____	____
13.	Troubleshoot fuel system malfunctions (truck, cargo, 2-1/2 ton, 6x6)	____	____	____
14.	Troubleshoot service brake malfunctions (truck, utility, 1/4 ton, 4x4)	____	____	____
15.	Replace starter	____	____	____
16.	Administer first aid to a nerve agent casualty (buddy-aid)	____	____	____
17.	Camouflage equipment	____	____	____
18.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
19.	Determine grid coordinates	____	____	____
20.	Determine magnetic azimuth using compass	____	____	____
21.	Load, reduce stoppage, clear M16A1 rifle	____	____	____
22.	Use challenge and password	____	____	____
23.	Put on and wear M17 series protective mask with hood	____	____	____

Selection and Classification Technical Area Working Papers

Arabian, J.M. (1984). Analysis of the relationship between SQT and aptitude area composite scores for Transportation Corps MOS. WP SC 84-18.

Arabian, J.M. (1984). Analysis of TRASANA written performance test scores as a function of aptitude area composite scores for Logistics MOS. WP SC 84-16.

Arabian, J.M. (1984). Conversion tables for estimating AFQT percentile distributions from aptitude area standard score distributions. WP SC 84-19.

Arabian, J.M. (1986). Standard setting: An annotated bibliography. WP SC 86-07.

Arabian, J.M., Wise, L.L., & Young, W. (1988). Army research to link standards for enlistment to on-the-job performance: Sixth Annual Report to Congress. WP SC 88-02.

Arabian, J.M., Young, W., & Wise, L. (1988). Army research to link standards for enlistment to on-the-job performance: Seventh Annual Report to Congress. WP SC 88-09.

Busciglio, H.H. (1989). The incremental validity of the figural reasoning and mazes tests in Project A's 1985 concurrent validation. WP SC 89-22.

Busciglio, H.H. (1990). The usefulness of Project A spatial tests for predicting comprehensive performance measures. (unnumbered)

Campbell, J.P. (1984). An outline of issues pertaining to EEO concerns and Project A. WP SC 84-24.

Chia, W.J., Owens-Kurtz, C., Peterson, N., and Szenas, P.L. (1988). Synthetic Validation Project: Pilot test results. WP SC 88-05.*

Czarnolewski, M.Y. (1988). Improving the selection of MOS 11H TOW gunners. WP SC 88-03.

Czarnolewski, M.Y., & Martin, C.J. (1988). Exploration of validities of Project A spatial tests. WP SC 88-08.

Eaton, N.K., Wing, H., & Mitchell, K. (1984). Alternate methods of estimating the dollar value of performance. WP SC 84-08.

Edwards, D.S., Claudy, J.J., Fischl, M.A., & Rumsey, M.G. (1984). Officer Selection Battery Forms 3 and 4 Manual for administration and scoring. WP SC 84-03.

Gast, I.F. (1988). Preliminary evaluation of two test-based strategies for identifying soldiers for the excellence in Armor Program. WP SC 88-04.

Grafton, F.C. (1988). Computing Army aptitude area composite scores (1980 metric) for Armed Services Vocational Aptitude Battery (ASVAB). Forms 6 and 7. WP SC 88-10.

Grafton, F.C., Czarnolewski, M.Y., & Smith, E.P. (1987). Relationship between Project A psychomotor and spatial tests and TOW2 Gunnery performance: A preliminary investigation. WP SC 87-10.

Grafton, F.C., Czarnolewski, M.Y., & Smith, E.P. (1989). Relationship between Project A psychomotor and spatial tests and TOW2 Gunnery performance: A preliminary investigation (Rev.). WP SC 89-01.

Grafton, F.C., Olson, D.M., Eaton, N.K., & Hanser, L.M. (1987). Project A concurrent validation results: Responses to CG TRADOC questions. WP SC 87-06.

Grafton, P.C., Olson, D.M., Walker, C.B., & Arabian, J.M. (1989). Current Army selection and classification research. WP SC 89-17.

Mitchell, K.J., & Hanser, L.M. (1984). ASVAB retest score gains. WP SC 84-10.

Mitchell, K.J., & Hanser, L.M. (1984). The 1980 youth population norms: Enlistment and occupational classification standards in the Army. WP SC 84-07.

Mitchell, K.J., Hanser, L.M., & Grafton, F.C. (1984). Status report on the implementation of the 1980 youth population norms for the Armed Services Vocational Aptitude Battery, Forms 11, 12, 13 and 14 and revision of the Army Clerical and Surveillance/Communications composites. WP SC 84-13.

Mitchell, K.J., & Harris, J.H. (1984). Army research to link standards for enlistment to on-the-job performance. WP SC 84-23.

Olson, D.M., & Hanser, L.M. (1983). Examination of current Infantry School accessions and projection of future needs. WP SC 83-02.

Olson, D.M., & Smith, E.P. (1987). Presentation of sex-related differences in spatial abilities and performance of Army soldiers for the JSSCWG. WP SC 87-03.

Olson, D.M., Walker, C.B., & Grafton, F.C. (1989). Project A: Current status. WP SC 89-18.

Olson, D.M., Walker, C.B., Rumsey, M.G., & Grafton, F.C. (1990). Project A: A summary. WP SC 90-01.

Osborne, W. (1984). A summary of the development procedures for job performance measurement in Project A. WP SC 84-20.

Park, R.K. (1988). Incorporating new predictors into the accelerated CAT-ASVAB project system. WP SC 88-06.

Rossmeissl, P.G., Martin, C.J., & Wing, H. (1983). Validity of ASVAB 8/9/10 as predictors of training success. WP SC 83-04.

Rossmeissl, P.G., & Wise, L.L. (1986). Project A data base security: An update and a reminder. WP SC 86-04.

Rumsey, M.G., Olson, D.M., & Knapp, D.J. (1989). Project A proponent feedback briefing reviewing status and results for MOS 63B: June 18, 1987. WP SC 89-03.

Silva, J.M. (1989). Usefulness of spatial and psychomotor testing for predicting TOW and UCOFT gunnery performance. WP SC 89-21.

Silva, J.M. (1989). Validity of AFQT and education level for predicting hands-on performance in 11H TOW heavy anti-tank missilemen. WP SC 89-07.

Silva, J.M., & Walker, C.B. (1989). Relationship of AFQT category to TOW gunnery performance. WP SC 89-04.

Smith, E.P., & Birnbaum, A.E. (1988). Data analysis to support the committee on the performance of military personnel. WP SC 88-07.

Walker, C.B. (ed.) (1989). ARI's nominations of new tests for the Technical Advisory Selection Panel (TASP). WP SC 89-06.

Walker, C.B. (1989). The process and results of predictor testing in TRADOC's Skills Selection and Sustainment (S3) Program. WP SC 89-23.

Young, W.Y., Austin, K.C., McHenry, J.J., & Wise, L.L. (1989). Concurrent Validation Codebook - 64C.- WP SC 89-16.

Young, W.Y., Austin, K.C., McHenry, J.J., & Wise, L.L. (1989). Sample Concurrent Validation Codebook. WP SC 89-02.

Working Paper -RS WP-84-18

ANALYSIS OF THE RELATIONSHIP BETWEEN SQT AND APTITUDE AREA COMPOSITE SCORES FOR TRANSPORTATION COPPS MOS

Jane M. Arabian
SELECTION AND CLASSIFICATION TECHNICAL AREA

June 1984

Reviewed by
Clinton B. Walker
Selection and Classification
Technical Area

Approved by
Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

p.6 374 0

Analysis of the Relationship Between SQT and
Aptitude Area Composite Scores for
Transportation Corps MOS

Jane M. Arabian

U.S. Army Research Institute
for the Behavioral and Social Sciences

The purpose of this research was to examine the relationship between the Armed Services Vocational Aptitude Battery (ASVAB) aptitude area (AA) composite scores and Skill Qualification Test (SQT) scores. The research was undertaken to provide the U.S. Army Transportation School with information about the job proficiency, given current AA entry standards, of soldiers in selected Transportation Corps military occupation specialities (MOS).

The nine AA composites currently used by the Army to select and classify enlisted personnel are created by combining ASVAB subtest scores. Aptitude area cut-off scores are used at the time of enlistment to determine eligibility for entering given MOS. Minimum AA scores are established for MOS to increase the likelihood that individuals allowed entry to an MOS will be able to perform their duties.

In order to ascertain how effectively an AA cut-off functions as an entrance standard, some measure of job performance is needed. The written SQT, which is now part of the Army Individual Training Evaluation Program (ITEP), is the best available measure of MOS-specific job knowledge. The Army has administered SQT to enlisted soldiers

since 1977, and has established the score of 60 points or better as acceptable (passing).

If the majority of soldiers with the minimum AA score in a particular MOS pass the SQT then the job proficiency of soldiers within that MOS could be seen as satisfactory. On the other hand, if the majority of soldiers meet the MOS AA cut-off but fail the SQT then the entrance standards may need to be changed. In such cases, it may be worthwhile to consider making entry into the MOS more difficult. Raising minimum AA scores would be one option for improving soldiers' proficiency levels in an MOS.

METHOD

Sample. U.S. Army Transportation School MOS were analyzed in this research. To permit meaningful analysis, each MOS was required to have at least 100 soldiers with both ASVAB and SQT data available. The two MOS meeting this requirement were: 57H (Cargo Specialist, n=183) and 64C (Motor Transport Operator, n=2038). A third MOS, 71N (Traffic Management Coordinator), was also analyzed even though the sample size was under 100 (n=92). Interpretation of the data for 71N should, therefore, be made with caution. All the soldiers in the three MOS entered the Army between October 1980 and September 1982.

Predictors. Each MOS in this research sample employed a different ASVAB AA composite as an entrance standard. Therefore, a different AA composite score was used as the predictor measure for each MOS. A minimum score of 85 on the General Maintenance (GM) AA composite

is required for entry into 57H. The GM composite is made up of scores from general science (GS), auto/shop information (AS), mathematics knowledge (MK), and electronics information (EI) ASVAB subtests. Soldiers entering 64C are required to have a score of at least 85 on the Operators/Food (OF) AA composite. The OF composite is formed by combining scores from verbal ability (VE), number operations (NO), auto/shop information (AS), and mechanical comprehension (MC) subtests. A minimum score of 95 on the Clerical (CL) composite (currently made up of verbal ability (VE), number operations (NO), and coding speed (CS) ASVAB subtests) is required for 71N.

Criteria. Scores obtained on the SQT by soldiers in the selected MOS served as the criterion measure. The SQT scores were obtained from tests administered during the first two quarters of the 1983 testing year.

Analysis. Two-way probability tables were calculated for AA composite and SQT scores. The composite score ranges were set at 5-point intervals, starting at the current cut-off score for each MOS, with a separate interval for all scores at or above 110. Five-point intervals were chosen as the basic AA category range because the Army's existing classification system establishes cut-off scores in 5-point intervals. Four SQT categories were created: below 60 (failing), greater than or equal to 60, 70, and 80.

RESULTS AND DISCUSSION

The data for 57H, 64C, and 71N are presented in Tables 1-3,

TABLE 1

Probability of Obtaining Given SQT Scores
by Aptitude Area Composite Score
(57H)

SQT Score Category	GM Aptitude Area Composite Score Category ≥ 85	≥ 90	≥ 95	≥ 100	≥ 105	≥ 110
≥ 80	15	14	16	14	27	38
≥ 70	45	44	47	43	45	50
≥ 60	80	82	87	81	82	88
< 60	20	18	13	19	18	12
Cumulative n	183	83	38	21	11	8

TABLE 2

Probability of Obtaining Given SQT Scores
by Aptitude Area Composite Score
(64C)

SQT Score Category	OF Aptitude Area Composite Score Category ≥ 85	> 90	≥ 95	≥ 100	≥ 105	≥ 110
≥ 80	59	64	69	72	77	80
≥ 70	90	92	93	95	95	96
≥ 60	98	98	98	99	99	99
< 60	2	2	2	1	1	1
Cumulative n	2038	1444	1037	700	453	319

TABLE 3

Probability of Obtaining Given SQT Scores
by Aptitude Area Composite Score
(71N)

SQT Score Category	CL Aptitude Area Composite Score Category $\geq$ 95	$\geq$100	$\geq$ 105	$\geq$ 110
$\geq$ 80	61	60	74	72
$\geq$ 70	80	79	89	88
$\geq$ 60	92	92	91	92
< 60	8	8	9	8
Cumulative n	92	62	35	25

respectively. For example, Table 1 indicates that for the 183 57H soldiers who have GM AA scores greater than or equal to 85, 80% obtained an SQT score of 60 or better. The probability of obtaining an SQT score of 60 or better was only 2% greater for soldiers having AA scores of 90 or more (82 minus 80).

It can be seen that, in general, as AA scores increase, the probability of obtaining a particular passing SQT score also increases for all the MOS in this research sample. This positive relationship between AA and SQT scores does not appear to be dramatic because of a "ceiling effect" (i.e., the fact that a vast majority of soldiers in each MOS with at least the lowest acceptable AA composite score pass the SQT). Anomolies in the data, such as a decrease in the probability of obtaining a given SQT score from one AA category to the next (higher) AA category, are most likely a result of the relatively small sample sizes.

Overall, given the high probability of obtaining an SQT score of 60 or better with an AA score at or above the entry standard (80%, 98%, and 92% for 57H, 64C, and 71N, respectively), the data would not support a recommendation to increase the AA cut-off scores. Furthermore, in the present sample 57H, which has the highest failure probability at the lowest end of the AA scores (20%), an increase of five points in the AA cut-off score would result in only a modest 2% increase in the probability of passing the SQT. This probable 2% performance increase would be achieved at the cost of eliminating from eligibility a large proportion of enlisted soldiers.

Based on the data obtained in this research, the cost of eliminating soldiers does not seem to be worth the benefit of a 2% improvement in performance as measured by the SQT.

Working Paper

RS-WP-84-16

ANALYSIS OF TRASANA WRITTEN PERFORMANCE TEST SCORES AS A FUNCTION OF APTITUDE AREA COMPOSITE SCORES FOR LOGISTICS MOS

Jane M. Arabian
SELECTION AND CLASSIFICATION TECHNICAL AREA

June 1984

Reviewed by
Clinton B. Walker CBW
Selection and Classification
Technical Area

Approved by
Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

pl8 6799

Analysis of TRASANA Written Performance Test Scores as a Function of Aptitude Area Composite Scores for Logistics MOS

Jane M. Arabian
Army Research Institute
for the Behavioral and Social Sciences

The U.S. Army Quartermaster School requested the Army Research Institute to examine the relationship between the Armed Services Vocational Aptitude Battery (ASVAB) aptitude area (AA) composite scores and soldier performance. Previous research (Rossmeissl & Eaton, 1984) reported on relationships between AA scores and Skill Qualification Test (SQT) performance in some Quartermaster Military Occupational Specialties (MOS): 76C, 76V, and 94B. With the cooperation of TRASANA, the research reported in this paper employed scores on written job skill and knowledge tests developed and administered by TRASANA as a measure of soldiers' job performance. The TRASANA test items were derived from soldier manuals and presented in a multiple choice format for each MOS. These tests were developed independently of the SQT which also measures a soldier's on-the-job performance and has been administered by the Army since 1977.

The purpose of examining the relationship between AA composite scores (predictor measure) and measures of job performance (criterion measure) is to obtain information for setting AA minimum (cut-off) scores. Aptitude area cut-off scores are used to determine eligibility for entry into MOS. The nine AA composites currently used by the Army to select and classify enlisted personnel are created by combining the ASVAB subtests' scores.

Aptitude area minimum scores are established for each MOS to increase the likelihood that individuals who are allowed entry to an MOS will be able to perform their jobs.

METHOD

Sample. Four U.S. Army Quartermaster School MOS were analyzed in this research. To permit meaningful analysis, each MOS was required to have at least 100 soldiers with both ASVAB and TRASANA test score data available. The four MOS meeting this requirement were: 76C (Equipment Records and Parts Specialist, n=476), 76P (Materiel Control Accounting Specialist, n=386), 76V (Materiel Storage and Handling Specialist, n= 570), and 76Y (Unit Supply Specialist, n=972). The data for this research came from soldiers in these MOS who were tested by TRASANA in 1983/1984.

Predictors. All the MOS in the research sample use the clerical (CL) composite as the standard for enlistment eligibility. Therefore, CL scores were used as the predictor variable in this research. The CL composite is currently formed by combining the verbal ability (VE), numerical operations (NO), and coding speed (CS) ASVAB subtests.

Criteria. The criterion measure for this research was the per cent correct score on the TRASANA written job skill and knowledge test. The criteria data were tabulated for soldiers who also had ASVAB data available.

Analyses. Since TRASANA was not allowed to release their raw data they prepared two-way frequency distribution tables of AA composite and written test scores for the Army Research Institute (ARI) to analyze. The TRASANA

tables were modified (i.e., cells were pooled) to conform more closely with previous ARI research. Frequency and cumulative percentage data were then derived from the revised tables. Composite score range was set at 5-point intervals, starting at the current cut-off score for each MOS, and with a separate category for all scores below the cut-off and at or above 110 points. Five-points was chosen as the basic AA category range because the Army's existing classification system establishes cut-off scores in 5-point intervals. Since there is no passing score established by the Army for the TRASANA tests, category ranges for these scores were set arbitrarily for convenience. With the exception of the lowest category (below 19% correct), category ranges were set at 15-point intervals.

RESULTS AND DISCUSSION

The data for each MOS are presented in probability tables (Tables 1-4 for 76C, 76P, 76V, and 76Y, respectively). Overall, there appears to be a positive relationship between AA scores and TRASANA written test performance in the four MOS. In general, as AA score catetories increase, test performance also tends to improve. Departures from this trend may be due to the relatively small sample sizes. With respect to the setting of AA cut-off standards, the data from the four MOS are somewhat difficult to interpret. The reason for this difficulty is discussed below.

Unlike SQT which has an established passing score of 60, the TRASANA test does not have a set passing grade. This is what primarily makes data interpretation difficult. Is 50% correct on the TRASANA test satisfactory

TABLE 1

Probability of Obtaining a Written Test Score
Given Aptitude Area Composite Score
(76C)

% Correct On TRASANA Written Test	CL Composite Score ≥ 95	≥ 100	> 105	≥ 110
≥ 80	.18	.24	.26	.27
≥ 65	.67	.69	.77	.79
≥ 50	.88	.88	.91	.95
≥ 35	.96	.96	.95	.97
≥ 20	.98	.98	.98	1.00
< 20	.02	.02	.02	.00
(n)	(302)	(208)	(129)	(77)

TABLE 2

Probability of Obtaining A Written Test Score
Given Aptitude Area Composite Score
(76P)

% Correct On TRASANA Written Test	CL Composite Score ≥ 90	≥ 95	≥ 100	≥ 105	≥ 110
≥ 80	.05	.08	.11	.15	.21
≥ 65	.28	.37	.41	.51	.62
≥ 50	.65	.73	.76	.80	.81
≥ 35	.92	.93	.94	.94	.92
≥ 20	.99	.99	1.00	1.00	1.00
< 20	.01	.01	.00	.00	.00
(n)	(301)	(157)	(111)	(74)	(53)

TABLE 3

Probability of Obtaining a Written Test Score
Given Aptitude Area Composite Score
(76V)

% Correct On TRASANA Written Test	CL Composite Score ≥ 90	≥ 95	≥ 100	≥ 105	≥ 110
≥ 80	.00	.00	.00	.00	.03
≥ 65	.06	.09	.12	.14	.18
≥ 50	.37	.46	.53	.54	.60
≥ 35	.76	.78	.79	.81	.82
≥ 20	.97	.98	.98	.97	.99
< 20	.03	.02	.02	.03	.01
(n)	(422)	(240)	(168)	(110)	(77)

TABLE 4

Probability of Obtaining a Written Test Score
Given Aptitude Area Composite Score
(76Y)

% Correct On TRASANA Written Test	CL Composite Score ≥ 95	≥ 100	≥ 105	≥ 110
≥ 80	.00	.00	.01	.00
≥ 65	.18	.20	.23	.26
≥ 50	.62	.66	.69	.74
≥ 35	.93	.94	.95	.94
≥ 20	1.00	1.00	1.00	1.00
< 20	.00	.00	.00	.00
(n)	(861)	(593)	(393)	(254)

or "passing"? What about 65% or 35% correct? If 50% correct is conservatively taken for a passing score then, of the total number of soldiers in each MOS with AA scores at or above the MOS-specific cut-off point, 88% of 76C, 65% of 76P, 37% of 76V, and 62% of 76Y pass their TRASANA written skill and knowledge test.

Looking at each MOS more closely, Table 1, for example, indicates that there is an 88% chance that a 76C soldier with an AA score at or above 95 will score 50% or better on the TRASANA written test. However, a soldier with an AA score at or above 100 is as likely as a soldier with an AA score at or above 95 to obtain at least 50% correct on the TRASANA test. In other words, for 76C, increasing the AA cut-off by five points does not improve the likelihood of soldiers scoring more than 50% on the skill and knowledge test. The data are similar for 76Y. Increasing the AA cut-off from 95 ot 100 only increases the chance of scoring 50% or better on the TRASANA test by 4% (from .62 to .66).

The probability data for 76P and 76V (Tables 2 and 3, respectively) indicate a different situation. Increasing the cut-off standard from 90 to 95 for these MOS improves the probability that soldiers would score at least at the 50% level on the skill and knowledge test. There is an 8% increase in probability for 76P and a 9% increase for 76V. (However, if a 65% passing score is accepted for the TRASANA test then raising the AA cut-off standard results in a 9% improvement in performance only for 76P.)

While these data indicate that 76P and 76V could possibly benefit from a 5-point increase in their respective AA entry standards, the

possible benefits must be weighed against the likely exclusion of a large number of enlisted soldiers from these two MOS. Other information about the MOS in this research should also be taken into account before policy decisions are made. For instance, Rossmeissl and Eaton included 76C and 76V in their research. That research, using SQT score as the criterion measure, suggested that increasing the AA cut-off for 76C but not 76V merited further consideration. The present research with the TRASANA tests suggests quite the opposite.

The differences between the results of the two research reports may be due to any one or more of the following factors: different soldiers in the MOS samples, differences between the content of the SQT and TRASANA tests, sample sizes, etc. A likely source of the difference in results is the difference in the content and scaling of the SQT and TRASANA tests. Problems of this nature will likely continue until standard performance measures in each MOS are developed and accepted for personnel decision-making.

References

Rossmeissl, P.G. and Eaton, N.K. (1984). An analysis of SQT scores as a function of aptitude area composite scores for Logistics MOS (Working Paper No. RS-WP-84-12). Alexandria, Virginia: Army Research Institute for the Behavioral and Social Sciences, Selection and Classification Technical Area.

Working Paper

RS-WP-84-19

CONVERSION TABLES FOR ESTIMATING AFQT PERCENTILE DISTRIBUTIONS FROM APTITUDE AREA STANDARD SCORE DISTRIBUTIONS

Jane M. Arabian
SELECTION AND CLASSIFICATION TECHNICAL AREA

July 1984

Reviewed by
Hilda Wing HW
Selection and Classification
Technical Area

Approved by
Newell K. Eaton E
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

1478 PP

Conversion Tables for Estimating AFQT Percentile Distributions from Aptitude Area Standard Score Distributions

Jane M. Arabian

U.S. Army Research Institute for the
Behavioral and Social Sciences
Alexandria, Virginia

Prior to entering the Army, individuals must take the Armed Services Vocational Aptitude Battery (ASVAB). Recruiting goals and selection standards are based on AFQT (Armed Forces Qualification Test) scores derived from the ASVAB. In addition to AFQT, nine aptitude area (AA) composite scores are also derived from the ASVAB. The AA scores are used by the Army to classify soldiers for entrance eligibility into a Military Occupational Specialty (MOS). While AFQT scores are used to set minimum trainability standards, AA scores are used to set specific trainability standards for each MOS. As such, AA scores are the most appropriate predictor measures to use when analyzing the distribution of quality among Army personnel where some

measure of job performance is the criterion variable. Nevertheless, in some situations an indication of the distribution of soldier quality in terms of AFQT percentile categories is needed. Since AFQT data are not universally available to Army analysts, the conversion tables were developed.

The attached tables, one for each aptitude area (AA) composite, are appropriate for use with AA scores from ASVAB 11/12/13/14 and other ASVAB AA scores which have been converted to the 1980 youth population norm group. The tables can be used to convert AA standard score distributions into expected AFQT category distributions. The conversion tables were calculated from cross tabulations (frequency distributions) of AA standard scores by AFQT percentile scores. The Profile of American Youth (NORC,1980) data set was used to derive the tables. The final NORC sample, based on 9,173 cases, contains ASVAB data from a representative sample of 18-23 year olds in the population; the corresponding population size (weighted sample) is over 25 million.

In order to use the tables, first obtain sample sizes for the 5-point aptitude area score ranges on the AA composite used for the MOS's entrance standard. Second, multiply the sample size of each AA score range by the percentages in the appropriate column for the AA score range. This results in the expected AFQT distribution for the individuals in the sample. For example, if 100 individuals scored between 90 and 94 on the CL AA composite

then you would expect 6.11 individuals (100 x .0611) to be in the AFQT I-IIIA category, 76.19 (100 x .7619) to be in the IIIB category, 17.70 (100 x .1770) to be in the IV category, and no one to fall in the V category.

A sample application of the conversion tables is also appended to illustrate further how the tables are used. The example is based on a fictitious set of data.

CONVERSION TABLES FOR ESTIMATING AFQT PERCENTILE DISTRIBUTIONS FROM APTITUDE AREA STANDARD SCORE DISTRIBUTIONS

CL Aptitude Area Composite

AA Standard Score

AFQT % ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.00	0.00	0.00	0.10	6.11	27.64	62.78	88.23	97.26	98.83	100.00	100.00	100.00			
IIIB	0.00	0.00	0.00	1.39	7.46	38.95	76.19	67.37	36.80	11.77	2.73	1.17	0.00	0.00	100.00			
IV	2.24	20.59	66.15	93.28	92.35	60.94	17.70	5.00	0.41	0.00	0.00	0.00	0.00	0.00	100.00			
V	97.76	79.40	33.85	5.33	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00			

CO Aptitude Area Composite

AA Standard Score

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.22	2.85	2.15	12.01	24.23	43.12	62.16	72.89	82.08	87.50	95.56	99.03	100.01	100.00	100.00	100.00
IIIB	0.00	1.61	7.49	12.01	27.96	36.16	44.00	38.52	25.18	21.16	14.57	11.76	4.44	0.97	0.00	0.00	0.00	0.00
IV	13.88	36.20	59.66	67.86	63.03	46.95	31.50	17.47	12.14	5.96	3.36	0.74	0.00	0.00	0.00	0.00	0.00	0.00
V	86.11	62.19	32.62	17.28	6.86	4.89	0.26	0.90	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

EL Aptitude Area Composite

AA Standard Scores

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.00	0.16	1.43	6.08	21.84	45.42	58.68	77.84	92.46	96.13	98.72	99.95	100.00	100.00	100.00	
IIIB	0.00	1.43	1.99	10.51	24.01	45.51	47.99	42.36	36.49	19.75	7.26	3.87	1.27	0.06	0.00	0.00	0.00	
IV	10.73	26.78	56.01	70.71	68.53	47.78	29.94	12.22	4.84	2.40	0.28	0.00	0.00	0.00	0.00	0.00	0.00	
V	89.27	71.79	42.01	18.62	6.03	0.64	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

FA Aptitude Area Composite

AA Standard Score

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.65	0.27	1.14	5.99	20.73	38.76	60.50	80.88	90.01	97.55	100.00	100.00	100.00	100.00	100.00	
IIIB	0.12	0.85	2.07	10.50	26.01	37.14	50.50	48.49	36.44	17.62	9.59	2.45	0.00	0.00	0.00	0.00	0.00	
IV	10.96	32.03	55.43	69.14	66.56	54.56	28.68	12.75	3.06	1.50	0.41	0.00	0.00	0.00	0.00	0.00	0.00	
V	88.91	67.12	41.85	20.09	6.29	2.30	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

GM Aptitude Area Composite

AA Standard Score

AFQT % ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.20	2.60	5.93	17.36	30.74	50.73	58.67	67.74	78.09	87.73	93.01	95.96	99.13	100.00	99.99	
IIIB	0.00	2.52	6.09	15.81	29.32	37.98	37.92	27.57	28.61	22.95	17.34	11.25	5.68	4.04	0.86	0.00	0.00	
IV	15.14	42.70	57.49	61.64	55.53	40.64	29.69	21.29	12.54	9.12	4.56	1.01	1.32	0.00	0.00	0.00	0.00	
V	84.86	54.78	36.21	19.94	9.21	4.02	1.64	0.42	0.18	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

MM Aptitude Area Composite

AA Standard Score

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.15	0.07	1.08	7.33	14.70	30.52	52.53	63.12	69.10	75.35	79.06	89.32	94.11	98.58	100.00	100.00	
IIIB	0.00	2.55	4.02	13.69	23.36	41.20	38.68	27.95	23.17	21.05	17.70	18.76	8.62	4.84	1.42	0.00	0.00	
IV	9.99	40.82	64.27	66.18	62.06	41.79	28.00	19.01	13.52	9.65	6.96	2.19	2.07	1.05	0.00	0.00	0.00	
V	90.00	56.47	31.64	19.06	7.26	2.31	2.81	0.50	0.18	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

OF Aptitude Area Composite

AA Standard Score

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.00	0.09	0.17	2.78	15.47	34.53	59.44	73.58	80.71	89.95	96.08	99.87	100.00	100.00		
IIIB	0.00	0.00	0.00	3.49	10.69	35.27	50.01	44.04	32.14	22.65	16.35	9.41	3.92	0.14	0.00	0.00		
IV	3.31	50.06	72.71	85.11	82.60	59.65	34.51	21.43	8.42	3.76	2.93	0.63	0.00	0.00	0.00	0.00		
V	96.68	49.93	27.29	11.31	6.54	2.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

SC Aptitude Area Composite

AA Standard Score

AFQT % ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.00	0.00	1.08	6.23	22.38	45.63	66.18	71.08	77.77	87.82	95.32	99.99	99.91	100.00		
IIIB	0.00	0.00	0.35	9.55	27.73	44.48	49.74	38.29	22.69	20.83	19.87	12.12	4.68	0.00	0.10	0.00		
IV	6.59	37.55	74.50	78.96	64.46	46.44	27.14	15.88	11.11	8.09	2.36	0.06	0.00	0.00	0.00	0.00		
V	93.40	62.45	25.14	11.49	6.73	2.87	0.76	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

ST Aptitude Area Composite

AA Standard Score

AFQT %ile	> 65	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149
I-IIIA	0.00	0.00	0.00	0.09	1.00	4.46	18.34	36.01	60.26	73.03	87.56	96.41	98.39	99.81	99.33	100.00		
IIIB	0.00	0.00	1.10	5.48	21.18	35.18	51.59	48.33	34.30	24.81	11.24	3.59	1.61	0.19	0.66	0.00		
IV	7.16	38.37	64.14	81.47	73.41	59.00	30.07	15.66	5.45	2.17	1.19	0.00	0.00	0.00	0.00	0.00		
V	92.83	61.63	34.76	12.95	4.41	1.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

SAMPLE APPLICATION OF THE AA TO AFQT EXPECTED DISTRIBUTION TABLES

Example

- School requires the following accessions in terms of CL scores:

	AA Score Range			
	95-99	100-104	105-109	110-114
n	487	380	220	103

- Using the n's and the AA to AFQT distribution table for the CL composite, the following is obtained:

expected AFQT distrib.	AA 95-99 n 487	100-104 380	105-109 220	110-114 103	Total	%
I-IIIA	135	239	194	100	668	56
IIIB	328	140	26	3	497	42
IV	24	2	0	0	26	2

NOTE: The numbers in the table have been rounded off to the nearest whole number.

86-07

Working Paper

RS-WP-~~07-86~~

STANDARD SETTING: AN ANNOTATED BIBLIOGRAPHY

Jane M. Arabian
SELECTION AND CLASSIFICATION TECHNICAL AREA

August 1986

SCTA FILE COPY

Clinton B. Walker

Reviewed by
Clinton Walker
Selection and Classification
Technical Area

Lawrence M Hanser

Approved by
Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

PB 9292

Standard Setting: An Annotated Bibliography

Jane M. Arabian[1]
U.S. Army Research Institute
Alexandria, Virginia

Introduction

The Army Research Institute is currently engaged in a large-scale, multi-year research project to improve the Army selection and classification system (Project A, "Improving the Selection, Classification and Utilization of Army Enlisted Personnel") and, thereby, increase the overall effectiveness of the force. The research is aimed at developing comprehensive selection and classification procedures to predict validly performance in Army training and occupational specialties.

A number of performance measures, including measures of training success, service-wide performance, and MOS-specific hands-on performance, were developed. The Army's rationale for developing multiple measures of job performance is based upon the knowledge that a soldier's job is multifaceted (i.e., many different kinds of tasks are involved) and there are multiple aspects to job performance (e.g., initiative, obedience, etc.). Therefore, in order to obtain information about the domain of job performance behaviors, the Army's research project has developed different kinds of tests to assess these different aspects of job performance. Composite

---

[1]The opinions, views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, expressed or implied, of the U.S. Army Research Institute for the Behavioral and Social Sciences or the Department of Defense or the United States Government.

scores, based on constructs derived from the performance measures, will be used as indices of job performance.

Preliminary analyses of field test data and other research lead to the expectation that data from the concurrent validation phase of the project will result in positive correlations between predictor and criterion measures of performance. This information, however, will not lead directly to the setting of enlistment standards. While it is possible to use cost trade-off models for selection and classification systems once a performance objective is determined, such models cannot identify the required performance objective. Some other method is needed to define performance requirements or standards before reasonable enlistment standards can be established.

To determine whether existing methods could be used to set job performance standards in the Army, the literature review in Appendix A was conducted. It is in the form of an annotated bibliography summarizing the content of each reference. While the bibliography is not intended to be exhaustive, it is representative of the published literature. Table 1 presents a listing of the bibliographic references and indicates the broad categories that reflect the content of each reference.

Overview of Standard Setting Procedures: General Issues

The majority of the references present definitions of standard setting and descriptions of various procedures that have been developed. The article by Glass (1978) describes a variety of standard setting methodologies. In addition, there are individual references for methodologies developed by Angoff (1971), Jaeger (1976, 1982, 1984) and Nedelsky (1954).

Table 1
Content Description of the
References

Reference	Issues/ Critique	Definitions	Description of Procedures	Comparison of Procedures	Education/ Certification	Job Performance Standards	Application (Specific)	"Utility"/ Decision Models
Andrew, et al (1967)				X	(X)		X	
Angoff, (1971)			X					
Block, (1978)	X							
Buck, (1977)	X	X	X					
Buck, (1975)		X						
Burton, (1978)		X	X					
Chuang, et al (1981)			X					X
DuBois, et al (1954)		X						
Eastman, (1981)						X	X	
Glass, (1978)	X	X	X					
Hambleton, (1978)		X			X			
Hambleton, et al (1980)		X	X		X			
Hambleton, et al (1979)			X					
Hofstee, (1983)	X		X					
Jaeger, (1982)			X	X	X			
Jaeger, (1976)	X	X						
Jaeger, et al (1984)				X	X		X	
Koffler, (1980)				X	(X)		X	
van der Linden, (1980)								X
Linn, (1978)	X							

Table 1
(Continued)

Reference	Issues/ Critique	Definitions	Description of Procedures	Comparison of Procedures	Education/ Certification	Job Performance Standards	Application (Specific)	"Utility"/ Decision Models
Livingston, et al (1983)				X	X		X	
MacPherson, (1981)			X			X	X	
Maslow, (1983)	X					X		
Messick, (1975)	X				X			
Nedelsky, (1954)			X					
Poggio, (1984)				X	X		X	
Popham, (1978)	X	X						
Reid, (1984)			X	X				
Scriven, (1978)	X					X		
Shepard, (1984)	X		X					
Shepard, (1983)	X		X					X
Shepard, (1976)	X	X						
Shikiar, et al (1985)	X					X		

The Poggio (1984) article provides a good comparison of various procedures used in Kansas to set standards on educational competency tests for reading and math. A large group of references deal with applications of standard setting procedures. By far, the bulk of the applications deal with setting educational standards (e.g., minimally acceptable levels of reading and math knowledge for high school graduates) and professional certification (e.g., minimal levels of knowledge that a grammar school teacher must possess to be certified or licensed to teach in a given state). There is clearly very little empirical, applied research dealing specifically with the determination of job performance standards for selection and classification purposes.

<u>Standard setting for certification vs. selection and classification</u>. A basic difference between standards for competence (mastery vs. nonmastery) or certification (CC) and for selection and classification (SC) is that the former essentially entail only one judgement. A CC standard is used to indicate that an individual meets the qualifications to be considered minimally competent.

Standard setting for SC purposes requires that two judgements be made. The first is a judgement of minimal competence or acceptable performance derived through measurement of job performance with job incumbents. In other words, a standard needs to be set on the criterion measure(s). A second cut-off score (standard) needs to be set on some predictor measure such that individuals who meet the standard on the predictor will be likely to meet, at some later point, the standard for acceptable job performance. Actually, two sets of these types of judgements may be needed: One

for selection and a second for classification. The first set of judgements, for selection, may require different considerations than the set of judgements for job classification. For example, selection standards may be based on concerns regarding supply and demand, trainability in a general sense and attrition, while classification standards would be based upon considerations of actual job performance. The Army model is based on this sort of multiple judgement approach.

Procedures: judgement and validity. Regardless of the context for setting standards, it should be understood that the application of any standard setting procedure requires judgement. While the judgements will be value-laden, they are not therefore arbitrary in the sense of being based solely on whimsey (cf. Hofstee, 1983). The judgemental nature of standard setting has been a focus of debate in the literature (e.g., Glass, 1978; Hambleton, 1978). However, rather than dismiss all standard setting procedures because they require judgements we need a more constructive approach. It seems reasonable to accept the fact that judgements are the basis of standard setting procedures and then examine the validity and impact of the resultant cut-off scores.

Unfortunately, the literature offers very little guidance for selecting one procedure over another based on considerations of validity. Indeed, it cannot be said that a test performance standard derived from any one procedure is intrinsically valid because of the particular procedure employed. Andrew and Hecht (1976) found that different groups of judges arrived at similar standards using any one procedure, but different standards were obtained when two different procedures were used by the same

groups of judges. According to Poggio's (1984) research, different procedures will consistently yield higher or lower standards. In fact, the validity of a cut-off score (i.e., the ability of the cut-off score to discriminate between minimally acceptable and unacceptable individuals) is likely to depend not only on the procedure used to set the cut-off but also on the content of the instrument(s) used to assess performance.

Norm-referenced vs criterion-referenced tests. It stands to reason that if a test does not suitably measure what it purports to measure then any standard or cut-off score based on that test will not be valid. This holds for tests developed within either a norm-referenced test (NRT) or criterion-referenced test (CRT) framework. Although a detailed description of NRT and CRT development procedures will not be presented here, several references in the bibliography (see Table 1, "Definitions") discuss the methods in more detail.

One point that bears emphasizing is the different goals of the tests. Basically, a NRT is designed to optimize discriminability between all individuals administered the test. A CRT is designed to maximize discriminability around the cut-off point for proficiency, and, technically, is composed only of items necessary for identifying proficiency in the content domain being tested. In terms of test item difficulty, NRTs tend to contain a range of item difficulties; CRT items, on the other hand, are viewed as homogeneous.

Despite different test development strategies, both types of tests will yield a distribution of response scores (e.g., number or percent correct). However, with a NRT one expects to obtain a more normal distri-

bution of scores while with a CRT one expects a somewhat skewed and peaked response distribution on, e.g., an end of course exam. It must be stressed that since ability is a continuous, not all-or-none, variable, test scores will always reflect a variety of abilities. Variations in scores are not simply measurement error on either a CRT or NRT. Therefore, scores on either type of test may be given a norm-referenced interpretation. Just as pass/fail standards are set on both NRT and CRT, one can discuss an individual's score in relation to all other scores from the exam with either a CRT or NRT.

With respect to the selection of a standard setting procedure, any procedure can be applied to either type (NRT or CRT) of test. However, the literature does imply that once a cut-off score or standard is set on a CRT, one may denote individuals whose scores fall above the standard as "masters" of the domain covered by the test. Individuals whose scores fall below the cut-off are designated as "non-masters" of the subject matter. In fact, one specific purpose of a standard on a CRT is to identify an individual as either a master or non-master of a particular skill domain (though not, necessarily, as more or less masterful than another individual). Appropriate labeling of individuals scoring above or below the cut-off point on an NRT is less clear. This may be attributed to the fact that since NRT items are sampled statistically (i.e., randomly from a pool) rather than on strict content domain grounds, it is less clear what an individual would be a master of, except the items on the test. The inference from the test items to the domain of skill is weaker for a NRT than for a CRT.

The report by Buck (1977) provides a good discussion of concerns and standard setting procedures for NRTs and CRTs. By way of summary, Buck states that "a test is not inherently norm-referenced or criterion-referenced. It is the manner in which a test is developed and interpreted that determines whether it is to be classified as norm- or criterion-referenced. It is conceivable that a test could be either norm- or criterion-referenced or both depending on the way in which it is developed, used and interpreted [p. 15]". Indeed, the Army's Project A measures encompass both NRT and CRT aspects. Due to the scope and purpose of the project, the development of the criterion (job) measures was based upon careful, comprehensive identification of the criterion domain followed by non-random sampling of tasks within the domain and construction of test items in such a way as to optimize discrimination of ability levels among individuals[1].

Modes of measurement. By and large, the published literature on standard setting deals with paper and pencil, multiple choice (recognition) tests. Applications of standard setting procedures to, for example, hands-on, rating, or interview assessment procedures are not represented in the literature (cf. Shikiar, et al 1985). This is not to say, however, that the existing procedures or the principles they embody cannot be made to accommodate different testing modes. Standard setting procedures can also be augmented to encompass the practice of using multiple tests and multiple test modes for criterion (job) performance measurement. Alterna-

---

[1] For a more detailed description of the criterion measures see: Campbell, C.H., Campbell, R.C., Rumsey, M.G., and Edwards, D.C. (1985). Development and field test of task-based MOS-specific criterion measures. Alexandria, VA: US Army Research Institute, in press.

tive approaches are described in Buck (1977).

Selecting and Applying Standard Setting Procedures: General Considerations.

Although the literature does not offer specific guidance on selecting one procedure over another for different situations, general recommendations or areas for consideration can be identified. Specific areas will be addressed in the following paragraphs: Acceptance of standards and modifying standards.

Acceptance of standards. One important consideration is the selection of the judges (standard setters). If representation of the end-users is included in the standard setting process, the likelihood that the resultant standards will reflect the interests, concerns, and needs of the users is increased. In the context of selection and classification in the military, it would be prudent to include individuals from the personnel, training, policy and field communities on standard setting panels. Involving several judges or groups of judges in the standard setting process may help to promote confidence in and acceptance of the standards. When independent groups of judges employ the same standard setting procedure and arrive at similar standards (cf. Andrew and Hecht, 1976), confidence in the standard will be increased.

Another consideration for increasing acceptance of standards is related to both the judges involved and the procedure selected. A procedure that seems convoluted to the judges or asks them to make decisions they do not feel qualified or knowledgeable enough to make is unsatisfactory. However, the very same procedure presented to a different group of judges

may meet with a more satisfactory response. Any procedure that causes judges to feel uneasy is not likely to result in a cut-off score that users will feel confident in implementing; the validity of the standard will be called into question. This is not to say that judges should be selected to "fit" the procedure. Rather, it is recommended that a procedure should be selected to "fit" the standard setters. Every effort should be made to ensure that that judges find the procedure credible and easy to apply.

Several sources have suggested providing judges with normative data so that their expectations of performance will not be unreasonable (e.g., Livingston and Zeiky, 1983; Shepard, 1976). Incorporation of normative data is likely to result in similar standards across judges which, in turn, is likely to improve confidence in the selected standard. Further, it should be noted that Jaeger et al. (1984) have found that iterative applications of a given standard setting procedure result in reduced variability across judges. Iterative applications, however, did not affect the mean recommended standard. The reduction in variability, i.e., better agreement among judges, is also likely to increase the confidence of the judges in the resultant standards.

Modifying standards. The preceeding discussion has concentrated on ways to maximize confidence in standards derived by any given procedure. It is important to ensure not only that the judges themselves are confident with the standard but also that the end-users and individuals directly affected by the use of the standards accept the results. Therefore, in addition to the above considerations, institutional requirements and

values must also be taken into account. No matter how confident judges may be in their decisions, if the standards appear too high or too low from an institutional perspective the standards will not be acceptable.

It was stated earlier that some procedures consistently yield higher standards than others. This means that some procedures will result in relatively high standards that are likely to produce false negative decisions, i.e., individuals will be classified as not minimally acceptable when, in fact, they would have been able to perform at a level acceptable to the organization. Conversely, procedures resulting in lower standards are more likely to produce some amount of false positive decisions. At an organizational level, then, consideration may be given as to whether false negative or false positive decisions are more serious or costly to the organization.

Decision theoretic and utility analyses can serve as a tool to "fine tune" a standard set by panels of judges. Decision theory is not a standard setting procedure; it is a technique for reducing the effects of measurement and sampling error (van der Linden, 1980). The goal of utility analysis is "to match the test dichotomy to the criterion dichotomy to ensure that the smallest number of classification errors will be made" (Shepard, 1983). One form of utility analysis is to determine the relative cost of one kind of error (false negative) against the cost of another kind of error (false positive). This may be a difficult, complex approach to apply especially with respect to deciding which cost factors should be used (e.g., cost of training, equipment loss, dollar value of performance).

Eaton et al. (1985)[1], have presented utility estimation techniques developed to be easier to apply in situations where "managers are more accustomed to considering the relative productivity of employees or crews than the costs of producing given levels of output ...[or]... where employees operate very complex, expensive equipment and/or are focal to the productivity of a costly system [p. 29]". The strategies presented by Eaton et al. consider changes in the number and performance level of system units for increased aggregate performance. As noted by the authors, these techniques still do not provide for easy linkage of performance quality to a single quantitative scale. The "linkage" maybe require complex judgements regarding the utility equivalence of different performance levels for different situations or groups of individuals.

Within Project A, attempts are being made to scale the value of different levels of performance. Utility scaling workshops will be conducted with military personnel. Their task will be to scale different performance levels of various Army occupations using the 50th percentile performance of the infantryman occupational speciality (11B) as a baseline. It is conceivable that the resultant scale value for the utility of individuals performing at the 90th percentile in some occupations will be lower than the scaled utility of the 50th percentile 11B.

Once the utility of performance levels is scaled onto a single dimension, information obtained from the scale may be used to modify test performance and/or entrance standards for different occupations in order to

----

[1] Eaton, N.K., Wing, H. and Mitchell, K. (1985). Alternative methods of estimating the dollar value of performance. <u>Personnel Psychology</u>, <u>38</u>, 27-40.

optimize selection and classification decisions. If, as expected, a given performance level (standard) does not have the same utility across occupational specialties, then the classification standards for each occupation may be modified in order to optimize the overall utility of the total enlisted force.

Conclusions

This review has started from the premise that tests on which standard setting procedures are applied have already been determined to be psychometrically sound. This is to say that the test must be valid, reliable, and follow the guidelines of the American Psychological Association[1] for test development practices. Once a test has been appropriately developed, a test performance standard (cut score) can be set.

Each standard setting procedure should be applied judiciously and care should be taken so that the mathematics involved in some of the procedures do not create a false sense of rigor. Every standard setting procedure is based on judgement. It is the responsibility of the developers, users, and overseers of the standard setting process to ensure that the judgements are sound, appropriate to and supportive of the goals and values of the organization or community served by the standards (cf. Hofstee, 1983). Further, it is incumbent on the responsible parties to evaluate and reevaluate the standards in terms of the impact the standards have on the organization. The basic objective of any standard is to help attain the

---

[1]American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC: Author.
American Psychological Association, Division of Industrial-Organization Psychology. (1980). Principles for the validation and use of personnel selection procedures. (Second edition) Berkeley, CA: Author.

critical goals and requirements of a given institution. Jobs, for example, change as do the needs of the organization. Accordingly, measures of job performance and standards based on those measures must be regularly appraised and modified as needed to ensure that the values of the organization are being met.

The concluding point of this review is that there cannot be one and only one correct standard. The notion that one correct standard can be determined for a given situation is logically inconsistent with the fact that performance or ability exists as a continuous variable. Any standard, no matter how it is derived, imposes an artifical dichotomy (e.g., pass vs fail, master vs non-master, etc). This not to suggest that standards should be eliminated or avoided. Standards do serve as useful tools in the selection and classification processes. Rather, the standard setting process cannot end with the determination of a particular standard. There is a need to continue evaluating the standard to ensure that the number and cost of the inevitable decision errors produced by that standard are minimized.

APPENDIX A

Andrew, B.J. and Hecht, J.T. (1976). A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 36, 45-50.

Abstract

Two standard setting procedures were employed by two groups of judges to set pass-fail levels for comparable samples of a nationally administered examination. These procedures were both designed to set standards in relation to the minimally qualified examinee. The study was undertaken to determine whether similar standards would be set for the same examination content when determined by different groups of judges, and whether the two procedures employed would result in similar standards for comparable samples of test content. In addition, the extent to which group consensus judgments might differ from individual judgments was also investigated. The results suggest that different groups of judges do set similar examination standards when using the same procedure, and that the average of individual judgments does not differ significantly from group consensus judgments. Significant differences were found, however, between the standards set by the two procedures employed. This finding was observed for both groups. The nature of these differences is described, and their implications for setting examination standards are discussed. (Author)

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational Measurement (pp. 508-600). Washington, DC: American Council on Education.

- Pg. 514-515: description of procedure:

Systematic procedure for deciding on minimum raw scores for passing and honors - think of "minimally acceptable person"

- go through test item by item;
- could such a person answer correctly the item under consideration: correct score 1
  incorrect score 0
- sum of scores = raw score [cut off] of minimally acceptable person
- have some number of independent judges decide by consensus without actually administering the tests
- results could later be compared with numbers and percent of examinees who actually earned the passing grades [validity studies: verify appropriateness of the initial cutting scores or correct them if necessary]
- or ask each judge to state the probability that the minimum acceptable person would answer each item correctly; the sum of the probabilities would represent the minimum acceptable score

- Pg. 531...suggestion of applying cut score procedure to Army.

Block, J.H. (1978). Standards and criteria: A response. Journal of Educational Measurement, 15, 291-295.

- Responds to Glass (1978) paper

- Argues that standard-setting techniques are not as arbitrary as Glass suggests

- Suggests developing new and better technique. ....promote broad-based humanistic procedure.

Buck, L.S. (1977). Guide to the setting of appropriate cutting scores for written tests: A summary of the concerns and procedures (Technical Memorandum 77-4). Washington, DC: Personnel Research and Development Center United States Civil Service Commission.

- Presents a review and summary of methods for establishing cut-scores

- Discusses issues as they apply to norm and criterion reference tests

- Pg. 9: summary of issues and models of test fairness
- Pg. 13-15: summary - cutting scores for NRTs
- Pg. 20-21: summary - cutting scores for CRTs
  (10 "models"... methodology different ... but no empirical or theoretical basis for selecting one over another)

- Pg. 13: "The test developer must realize that the process of setting a cutting score cannot be totally analytic, as it is impossible to assume a purely objective attitude".

Buck, L.S. (1975). Use of criterion-referenced tests in personnel selection: A summary status report (Technical Memorandum 75-6). Washington, DC: Personnel Research and Development Center United States Civil Service Commission.

- Discusses validity of CRTs and measures of reliability (actually does little more than reference papers dealing with the topics)

- See pg. 21 ... measures of reliability.
  pg. 23 ... validity
  pg. 26 note ... reliability/correlation estimate

- Provides 28-page annotated bibliography

Burton, N.W. (1978). Societal standards. Journal of Educational Measurement, 15, 263-272.

- Pg. 264 - definition of criterion - differences in emphasis:
  1) criterion [variable] - trait to be measured (traditional definition)
  2) - specification of minimum levels of performance (Glaser and CRTs)

- Pg. 266 - 3 types of methods:
  1) standards based on theories (learning hierarchies)
  2) standards based on expert consensus
  3) standards based on practical necessities (minimal competencies for real-life)

Chuang, D.T., Chen, J.J., & Novick, M.R. (1981). Theory and practice for the use of cut-scores for personnel decisions. _Journal of Education Statistics_, _6_, 129-152.

- Mathematical model

- Optimize utility ... final cut-score set by utility function...
  ... assumes some cut score has already been determined...

- Does not specify how cut scores are set

DuBois, P.H., Teel, K.S., & Petersen, R.L. (1954). On the validity of proficiency tests. Educational and Psychological Measurement, 14, 605-616.

- A proficiency test is considered valid if it discriminates between the proficient and non-proficient in a given skill... "while an aptitude test may have an indefinite number of validities, depending on the criteria which it predicts with varying degrees of success, a valid proficiency test must measure what it purports to measure. No other concept of validity is applicable, the only variation possible is in the method used in arriving at the estimate of validity" (pg. 605)

- Coverage and discrimination power are independent dimensions of a proficiency test

- Difficulty analysis - item has p-value of .50 (passed by 50% of the group)... its SD is at maximum and makes maximum numbers of discriminations... A range of item difficulties, from very easy to very hard with a mean at about .50 is optimal for differentiating within a given population

- Types of validity:
  - Validity by "Direct Judgment" -- uses SMEs
  - Work sample validity -- correlation with work sample that is representative and meaningfully measures the skill
  - Class validity -- e.g. high vs. medium vs. low proficiency
  - Curricular validity - untrained vs. trained

Eastman, R.F. (1981). Supervisor ratings as criteria for Skill Qualification Tests. In S.F. Bolin (Chair), Panel on skill qualification testing: An evolving system (pp. 1356-1366). Arlington, VA: 23rd Annual Conference of the Military Testing Association. [DTIC, ADP 001400]

- Correlation between supervisor ratings of overall job performance and SQT scores for 67N (r=.74)

- Used ratings as criteria to determine optimum cut score for performers vs. non-performers

  ratings:
  - competent/not competent/don't know
  - cross out name if you don't supervise the individual
  - use a "+" sign to indicate one of the best soldiers and a "-" sign to indicate one of the poorest soldiers

- Each soldier rated by 3-6 supervisors

- Plot (tabulate) distribution of performers (rating $\geq$ 3) by SQT score

- Findings: SQT cut-score could be lowered and be more consistent with perceptions of supervisory personnel; it may be better to have supervisor rank-order soldiers instead of having to designate a soldier as a non-performer.

Glass, G.V. (1978). Standards and criteria. Journal of Educational Measurement, 15, 237-262.

- Pg.243: Criterion: the definition has been corrupted; originally referred to a criterion-referenced test ... meaning a scale of behavior linked to a test scale ... now criterion taken to be synonymous with "standard" or "cut score"

- Glass expresses strong concern regarding the arbitrary nature of setting a standard and the notions of a standard

- Six classes of methods for establishing cut-offs:

1. Performance of others - reference parameters of existing population of examinees - e.g., median score, 50th percentile ... essentially normative... not "behaviorally informative"; criterion-reference test theorists would find this approach to be an inappropriate method.
2. "Counting Backwards from 100%" - given the nature of criterion-references test (objectives)...would expect perfect scores...but allowances must be made for, e.g., measurement error and clerical mistakes...but how much??...the method is highly judgmental and too vague
3. Bootstrapping on other criterion scores - use other determinations of competence to select a group then match the group against the score distribution of some other test...problems 1) the 2 tests must be correlated, but it will never be perfect, therefore, you will make, e.g., false positive & false negative decisions, and you still have the problem of .. how was the standard set on the first test...circularity problem
4. Judging minimal competence - study a test and determine the required score for a minimally competent individual (cf. Nedelsky; Ebel)...see pg. 246 for Nedelsky method; pg 247 for Ebel method; pg. 248 for Angoff method; problems: 1) consistency and reliability of judges; 2) logical psychological status of concept of minimal competence.
5. Decision theoretic approaches - cutoff on an external criterion assumed as a "given"...vary score on the criterion-referenced test to say, minimize false negatives...approach simply postpones decision regarding the setting of a cut off... still "arbitrary"
6. "Operations Research" Methods - based on OR approach of maximizing a valued commodity by finding an optimum point on a mathematical curve or graph--- must have a non-monotonic curve...could have composite with a second valued outcome; but then have the problem of how to weight the composite...or look for the point of diminishing returns...(no further gain) how do you decide non-arbitrarily

- Pg. 258... standard-setting procedure may involve more precision than the test itself has... no matter what procedure used, there is still the element of the arbitrary

- Glass favors a comparative approach, e.g., improvement (change in performance) but you still have the questions re: how much change is good/-sufficient...how much loss before action should be taken... same problem as with criterion score, but he claims one has still gained clarity and consensus even if all problems were not solved.

Hambleton, R.K. (1978). On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 15, 277-290.

- Validity of cut score depends on how accurately it separates examinees into mastery states...usually the criterion is some external measure of performance or instructed vs. non-instructed groups

- Methods are based on the consideration of item content, educational consequences, psychological and financial costs, performance of others, errors due to guessing and item sampling...all arbitrary

- Against Glass' (1978) recommendation of using change scores.

Hambleton, R.K. & Eignor, D.R. (1980). Competency test development, validation, and standard setting. In R.M. Jaeger & C.K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences (pp. 367-396). Berkeley, CA: McCutchan Publishing Corporation.

- Focuses on making competency judgment for individuals - not groups (e.g....program evaluation)

- "A minimum competency test is designed to determine whether an examinee has reached a prespecified level of performance necessary to each competency being measured"..."standard" [or "cutoff score" or "minimal proficiency level"] is a point on a test score scale which is used to separate examinees into two categories"...master/nonmaster - a standard is set for each competency measured by a test..."competency tests are a special type of criterion-referenced test" - requires "information about levels of individual performance relative to well-defined content domains (referred to as "domain specifications)"

- 4 important topics: 1) improved guidelines for preparing domain specifications
  2) guidelines for evaluating competency tests and test manuals
  3) research on the relationship among test length, test score reliability, and test score validity
  4) consideration of issues and methods for determining standards, as well as guidelines for implementing each method

- Goes through 12-step model for developing and validating competency tests

- Pg. 377: test length formula for criterion reference tests, (vs. Spearman-Brown for norm-referenced tests)

- Continuum vs. state models (all-or-none): in the latter, test true-score performance is viewed as all-or-none, true-score standard is set at 100%, after consideration of measurement error the observed-score standard is set at a value less than 100%...use normative information as an aid in making decisions (in case experience of judges may have been with unusual students)

*See pg. 383-384...different models' use of (need for) utility values
pg. 386 - comparison of standard setting procedures
pg. 392 - for empirical methodology and the need for external criterion measures (see refs.)
pg. 3 - latent trait models...feasibility with competency tests??, equating scores from one form of competency test to another.

PB 8877
~~[illegible]~~

Selection and Classification Technical Area Working Paper RS-WP-88-2

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE:
6TH ANNUAL REPORT TO CONGRESS

Jane M. Arabian
US Army Research Institute

Lauress L. Wise and Winnie Young
American Institutes for Research

February 1988

REVIEWED BY:
Darlene M. Olson
Darlene M. Olson
Research Psychologist

APPROVED BY:
Lawrence M. Hanser
Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Cleared by: David W. Witt
For Newell K.Eaton, Director

Manpower and Personnel Research Laboratory

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

CHAPTER 3

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE

CURRENT STATUS AND ACCOMPLISHMENTS

The Army Research Institute is currently engaged in a large-scale, multi-year project to improve the Army selection and classification system and, thereby, increase the overall effectiveness of the force. The goal of the Army's program for increasing the efficiency of enlisted personnel selection and utilization is to enable the Army to meet its peacetime and mobilization missions through improved matching of individuals to Military Occupational Specialties (MOS). The research is aimed at developing comprehensive selection and classification procedures to validly predict performance in Army training and occupational specialties.

A complete description of the Army's research program and accomplishments have been published separately by the U.S. Army Research Institute for the Behavioral and Social Sciences in the annual reports for Project A, "Improving the Selection, Classification and Utilization of Army Enlisted Personnel" (1983, 1984, 1985, 1986). A detailed description of the project also appears in previous Annual Reports to Congress on Joint-Service Efforts to Link Standards of Enlistment to On-the-Job Performance (1983, 1984, 1985, 1986).

The Army's research focuses on 19 MOS. The MOS were selected to be representative of the Army and include all operational ASVAB aptitude area composites. Blacks, Whites, Hispanics, males, and females are present in these MOS in the same proportions as in total accessions. These MOS represent 44% of annual Army enlistments.

A number of performance measures, including measures of training success, service-wide performance, and MOS-specific hands-on performance, were developed for these MOS. For reasons of cost efficiency, not all measures were developed for all 19 MOS. All project criterion measures were developed for the following MOS and serve as the focus of this report:

1. 11B - Infantryman
2. 13B - Cannon Crewman
3. 19E - Tank Crewman
4. 31C - Radio Teletype Operator
5. 63B - Light Wheel Vehicle/Power Generation Mechanic
6. 64C - Motor Transport Operator
7. 71L - Administrative Specialist
8. 91A - Medical Specialist
9. 95B - Military Police

Measures of training success and service-wide performance were developed for the following specialties:

1. 12B - Combat Engineer
2. 16S - MANPADS Crewman
3. 27E - TOW/Dragon Repairman
4. 51B - Carpentry/Masonry Specialist
5. 54E - Chemical Operations Specialist
6. 55B - Ammunition Specialist
7. 67N - Utility Helicopter Repairman
8. 76W - Petroleum Supply Specialist
9. 76Y - Supply Specialist
10. 94B - Food Service Specialist

A complete report of accomplishments and current status of the Army's research to link enlistment standards to job performance has been published by the Army Research Institute (1983, 1984, 1985, 1986) and is available for distribution. What follows includes a summary of the Army's accomplishments in the area of job performance measurement.

As reported last year in the Annual Report to Congress the Army has completed the concurrent validation phase, i.e., joint administration of predictor and criterion measures, of its project. The research not only established the feasibility of measuring job performance but also identified five performance constructs, described later in this chapter, thereby providing support for the notion that job performance is multi-dimensional. In addition, the research established that there is a positive relationship between ASVAB and job performance during the first term of service. More specifically, the higher an applicant's ASVAB score the better he or she performs the technical or "can-do" aspects of the job. The correlations ranged from .23 to .68. The will-do or motivational factors are not well-predicted by ASVAB, but the Army's new predictor measures are expected to improve the prediction of the non-cognitive job performance dimensions.

The Army is currently engaged in the longitudinal validation phase of its project. In the past year, fifty-three thousand Army entrants were administered the Army project's predictor battery at eight reception battalions. Collection of training (school) knowledge measures, administered to soldiers in the sample at the end of Advanced Individual Training (AIT), has begun and will end in May of 1988. Administration of the job performance measures is scheduled to be conducted between July and November 1988. The measures are currently being revised and pilot tested; they will be field tested between January - March.

Implementation of the project's findings currently includes three efforts. First, with respect to the use of ASVAB, modifications have been made to the aptitute area composites to improve can-do assignment efficiency. Forty-four MOS are in the process of being reassigned to different entry composites in

order to: increase validity, increase consistency of composite use within a career management field, and increase the accuracy of prediction for women and blacks. The second major implementation effort involves one of the new predictors, Assessment of Background and Life Experiences (ABLE), a temperament/personal history instrument. The Army is presently planning for an Initial Operational Test and Evaluation (IOT&E) of ABLE. ABLE will help the Army to predict better the job dimensions of: Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. The third implementation is designed to better capitalize on individual psychomotor and spatial attributes in MOS which require them for improved gunnery performance.

RESEARCH DESIGN

Types of Measures Developed

Hands-on performance measures, job knowledge tests, and performance rating scales were developed for training success, service-wide performance, and MOS-specific performance for the MOS listed previously. The Army's rationale for the development of multiple measures of job performance is based upon the knowledge that a soldier's job is multi-faceted and there are different aspects of job performance. Therefore, the Army's research project has developed different kinds of tests to assess these different aspects of job performance and thereby obtain information about the domain of job performance behaviors. A more complete description of these measures appears in the Project A Annual Reports (1983, 1984, 1985, 1986) and in previous Annual Reports to Congress (1983, 1984, 1985, 1986).

Data Collection Procedures

Data collection began with the briefing of local military commanders, examination of the test sites, equipment and supplies, training of test administrators and scorers, and a dry run of testing procedures. Test site managers were appointed to supervise the actual data collections and were responsible for controlling the quality and flow of data from the testing.

The concurrent validation data collection began in June and ended in November 1985. As reported in the Fourth Annual Report to Congress, data were collected at fourteen different sites. The data from 5200 soldiers in 9 MOS, who were tested on all the project criterion measures, were entered into the research data base. Data from an additional 4000 soldiers in the 10 other MOS, who were tested only with the measures of training success and service-wide performance were also entered. Analyses began as soon as all the data were collected and entered in the longitudinal research data base.

RESULTS OF DATA ANALYSES

Common Data Analyses

In order to provide comparable data across services, the Principal Deputy Assistant Secretary of Defense for Force Management and Personnel directed the following analyses be conducted by each service. The results of the analyses are presented below.

Sample Description. The analyses are based on data collected during the Army's concurrent validation data collection in 1985. Soldiers were in Skill Level One, entry-level positions in their respective MOS primarily with 18-24 months experience in the Army.

Table 1 presents descriptive statistics for the variables included in these analyses. It should be noted that the hands-on performance test (HOPT) scores have been converted to standard T-scores in order to allow comparison across MOS. Time in service is the number of months, averaged across soldiers, of Army service. In accordance with the Army's sampling plan, the means are between 18 and 24 months, with small variances (standard deviations). Average AFQT percentile score for each MOS sample is also presented, along with the number of soldiers classified as high school diploma graduates (HSDG) or non-high school diploma graduates (NHSDG).

Table 2 also presents descriptive statistics for the sample, this time broken into four groups for time in service and two AFQT groups, upper 50th percentile (I-IIIA) and lower 50th percentile (IIIB-IV). Again, given the design of the Army's project, few soldiers were expected in the 1-12, 25-36, and 37+ Time in Service groups.

Reliability Results. Two indices of reliability, split-half and coefficient alpha, for HOPT were computed and are presented in Table 3. Both uncorrected and corrected reliability estimates are included along with the standard error of measurement. The table also includes the number of tasks in the HOPT scores for these common data analyses.

Hands-On Performance Test, Aptitude, Education, and Job Experience Relationships. Intercorrelations among the measures are presented for each MOS. Uncorrected correlations are depicted in Table 4. The correlations in Table 5 have been corrected for range restriction using a univariate correction procedure as agreed upon by the Joint Service Job Performance Measurement Working Group for the purposes of these analyses.

Service - Specific Data Analyses

We believe the most appropriate data showing the link between

Table 1

Descriptive Statistics for HOPT, Time in Service, Aptitude, and Educational Attainment

MOS		HOPT (Standard T-Score)	Time in Service (Months)	AFQT (Percentile)	Educational Attainment (N)	
	Mean	50	20.4	56.2	423	HSDG
11B	SD	10	5.6	21.8	55	NHSDG
	N	491	478	491		
	Mean	50	21.8	47.0	353	HSDG
13B	SD	10	6.1	19.7	73	NHSDG
	N	434	425	464		
	Mean	50	19.4	53.9	317	HSDG
19E	SD	10	4.6	21.5	50	NHSDG
	N	378	367	394		
	Mean	50	20.5	58.9	245	HSDG
31C	SD	10	5.6	19.2	29	NHSDG
	N	275	274	289		
	Mean	50	19.9	49.7	397	HSDG
63B	SD	10	5.0	19.3	64	NHSDG
	N	443	461	478		
	Mean	50	20.0	43.1	420	HSDG
64C	SD	10	4.6	18.3	66	NHSDG
	N	481	486	507		
	Mean	50	22.3	56.5	401	HSDG
71L	SD	10	4.5	19.2	7	NHSDG
	N	416	408	427		
	Mean	50	20.6	59.1	310	HSDG
91A	SD	10	5.4	18.3	68	NHSDG
	N	392	378	392		
	Mean	50	21.3	62.1	525	HSDG
95B	SD	10	5.0	17.3	44	NHSDG
	N	577	569	597		

Table 2

Mean HOPT Scores (Standardized T-Scores) by Aptitude and Job Experience Levels for 9 Jobs

Time in Service (Months)		MOS 11B AFQT I-IIIA	MOS 11B AFQT IIIB-IV	MOS 13B AFQT I-IIIA	MOS 13B AFQT IIIB-IV	MOS 19E AFQT I-IIIA	MOS 19E AFQT IIIB-IV
1-12	Mean	52.8	45.2	41.2	46.7	47.0	37.6
	SD	9.0	10.4	--	11.1	11.8	8.5
	N	6	7	1	5	7	4
13-24	Mean	51.9	48.0	49.8	48.1	51.0	47.7
	SD	9.8	10.3	10.4	9.3	9.6	10.8
	N	213	155	108	158	150	134
25-36	Mean	50.8	48.3	53.6	52.3	55.4	52.3
	SD	9.2	9.5	10.4	9.8	6.6	9.2
	N	41	47	38	74	25	32
37+	Mean	49.5	41.8	54.5	53.8	--	51.7
	SD	8.4	15.2	11.6	7.6	--	--
	N	4	5	5	6	--	1
Total Group	Mean	51.7	47.8	50.8	49.5	51.4	48.3
	SD	9.6	10.2	10.5	9.6	9.5	10.7
	N	264	214	152	243	182	171

Time in Service (Months)		MOS 31C AFQT I-IIIA	MOS 31C AFQT IIIB-IV	MOS 63B AFQT I-IIIA	MOS 63B AFQT IIIB-IV	MOS 64C AFQT I-IIIA	MOS 64C AFQT IIIB-IV
1-12	Mean	49.2	46.0	40.7	50.6	54.9	52.6
	SD	9.7	8.8	9.1	3.5	11.3	8.3
	N	8	7	6	7	5	6
13-24	Mean	52.0	47.2	51.3	48.9	51.3	48.6
	SD	9.9	9.2	10.4	9.5	10.0	9.8
	N	140	74	156	193	96	258
25-36	Mean	50.2	46.2	50.0	51.9	55.0	50.1
	SD	10.6	12.9	10.3	10.1	8.9	10.7
	N	16	10	15	47	32	64
37+	Mean	51.9	42.5	25.1	57.9	--	--
	SD	16.7	9.0	--	7.7	--	--
	N	2	3	1	3	--	--
Total Group	Mean	51.7	46.8	50.7	49.6	52.4	49.0
	SD	10.0	9.5	10.6	9.6	9.9	9.9
	N	166	94	178	250	133	328

Table 2 (Continued)

Mean HOPT Scores (Standardized T-Scores) by Aptitude and Job Experience Levels for 9 Jobs

Time in Service (Months)		MOS 71L AFQT I-IIIA	MOS 71L AFQT IIIB-IV	MOS 91A AFQT I-IIIA	MOS 91A AFQT IIIB-IV	MOS 95B AFQT I-IIIA	MOS 95B AFQT IIIB-IV
	Mean	48.9	53.5	49.8	47.3	49.5	47.5
1-12	SD	15.4	13.5	11.1	10.2	9.0	8.0
	N	6	4	10	4	18	6
	Mean	53.2	46.5	51.3	47.2	50.6	47.3
13-24	SD	9.8	9.4	9.8	10.0	9.7	10.5
	N	179	119	195	88	290	111
	Mean	51.4	47.8	51.2	48.1	52.3	49.7
25-36	SD	8.8	8.9	9.8	10.6	10.1	10.2
	N	49	38	42	36	83	41
	Mean	48.5	44.7	57.0	50.3	48.0	--
37+	SD	3.8	--	3.2	--	--	--
	N	2	1	2	1	1	--
	Mean	52.7	47.0	51.3	47.5	50.9	47.9
Total	SD	9.7	9.3	9.8	10.1	9.7	10.3
Group	N	236	162	249	129	392	158

Table 3

Reliability Estimates and Standard Errors of Measurement for Job-Specific Hands-on Performance Tests

MOS	Number of Tasks		Split-Half	Coefficient Alpha	SEM[a]
11B	14	ucoef.[b]	.56	.55	6.71
		ccoef.[c]	.57	.57	
13B	17	ucoef.	.76	.70	5.52
		ccoef.	.76	.70	
19E	15	ucoef.	.57	.56	6.66
		ccoef.	.59	.57	
31C	15	ucoef.	.81	.69	5.58
		ccoef.	.83	.72	
63B	15	ucoef.	.47	.40	7.74
		ccoef.	.47	.40	
64C	16	ucoef.	.66	.60	6.32
		ccoef.	.68	.63	
71L	15	ucoef.	.73	.68	5.65
		ccoef.	.76	.72	
91A	17	ucoef.	.59	.72	5.31
		ccoef.	.62	.74	
95B	16	ucoef.	.63	.52	6.90
		ccoef.	.65	.55	

[a]Standard Error of Measurement calculated with the Coefficient Alpha reliability estimates and HOPT standard T-Score standard deviations.

[b]Uncorrected reliability estimate.

[c]Corrected reliability estimate.

Table 4
Uncorrected Intercorrelations

Intercorrelations Between Measures for MOS 11B

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.06	.23	-.12
2 Time in Service	-.06	1.00	-.04	-.05
3 AFQT Percentile	.23	-.04	1.00	-.06
4 Educational Attainment	-.12	-.05	-.06	1.00

Intercorrelations Between Measures for MOS 13B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.18	.11	-.02
2 Time in Service	.18	1.00	-.01	-.04
3 AFQT Percentile	.11	-.01	1.00	.15
4 Educational Attainment	-.02	-.04	.15	1.00

Intercorrelations Between Measures for MOS 19E

Measure	Measure 1	2	3	4
1 HOPT	1.00	.26	.22	.08
2 Time in Service	.26	1.00	-.01	.09
3 AFQT Percentile	.22	-.01	1.00	.01
4 Educational Attainment	.08	.09	.01	1.00

Intercorrelations Between Measures for MOS 31C

Measure	Measure 1	2	3	4
1 HOPT	1.00	.01	.32	-.04
2 Time in Service	.01	1.00	.03	.03
3 AFQT Percentile	.32	.03	1.00	.03
4 Educational Attainment	-.04	.03	.03	1.00

Intercorrelations Between Measures for MOS 63B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.07	.07	.05
2 Time in Service	.07	1.00	-.10	-.13
3 AFQT Percentile	.07	-.10	1.00	.09
4 Educational Attainment	.05	-.13	.09	1.00

Table 4 (Continued)
Uncorrected Intercorrelations

Intercorrelations Between Measures for MOS 64C

Measure	1	2	3	4
1 HOPT	1.00	.04	.21	-.06
2 Time in Service	.04	1.00	.08	.14
3 AFQT Percentile	.21	.08	1.00	.11
4 Educational Attainment	-.06	.14	.11	1.00

Intercorrelations Between Measures for MOS 71L

Measure	1	2	3	4
1 HOPT	1.00	-.02	.33	-.06
2 Time in Service	-.02	1.00	-.08	.05
3 AFQT Percentile	.33	-.08	1.00	-.07
4 Educational Attainment	-.06	.05	-.07	1.00

Intercorrelations Between Measures for MOS 91A

Measure	1	2	3	4
1 HOPT	1.00	.05	.25	-.05
2 Time in Service	.05	1.00	-.10	.28
3 AFQT Percentile	.25	-.10	1.00	.02
4 Educational Attainment	-.05	.28	.02	1.00

Intercorrelations Between Measures for MOS 95B

Measure	1	2	3	4
1 HOPT	1.00	.06	.19	-.10
2 Time in Service	.06	1.00	.00	.06
3 AFQT Percentile	.19	.00	1.00	-.03
4 Educational Attainment	-.10	.06	-.03	1.00

Table 5
Intercorrelations Corrected for Range Restriction

Intercorrelations Between Measures for MOS 11B

Measure	1	2	3	4
1 HOPT	1.00	-.07	.30	-.13
2 Time in Service	-.07	1.00	-.05	-.05
3 AFQT Percentile	.30	-.05	1.00	-.08
4 Educational Attainment	-.13	-.05	-.08	1.00

Intercorrelations Between Measures for MOS 13B

Measure	1	2	3	4
1 HOPT	1.00	.17	.16	.00
2 Time in Service	.17	1.00	-.01	-.04
3 AFQT Percentile	.16	-.01	1.00	.21
4 Educational Attainment	.00	-.04	.21	1.00

Intercorrelations Between Measures for MOS 19E

Measure	1	2	3	4
1 HOPT	1.00	.26	.29	.08
2 Time in Service	.26	1.00	-.02	.09
3 AFQT Percentile	.29	-.02	1.00	.02
4 Educational Attainment	.08	.09	.02	1.00

Intercorrelations Between Measures for MOS 31C

Measure	1	2	3	4
1 HOPT	1.00	.02	.45	-.03
2 Time in Service	.02	1.00	.05	.03
3 AFQT Percentile	.45	.05	1.00	.05
4 Educational Attainment	-.03	.03	.05	1.00

Intercorrelations Between Measures for MOS 63B

Measure	1	2	3	4
1 HOPT	1.00	.06	.11	.06
2 Time in Service	.06	1.00	-.15	-.14
3 AFQT Percentile	.11	-.15	1.00	.13
4 Educational Attainment	.06	-.14	.13	1.00

Table 5 (Continued)
Intercorrelations Corrected for Range Restriction

Intercorrelations Between Measures for MOS 64C

Measure	Measure 1	2	3	4
1 HOPT	1.00	.06	.33	-.02
2 Time in Service	.06	1.00	.13	.16
3 AFQT Percentile	.33	.13	1.00	.17
4 Educational Attainment	-.02	.16	.17	1.00

Intercorrelations Between Measures for MOS 71L

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.05	.46	-.09
2 Time in Service	-.05	1.00	-.12	.05
3 AFQT Percentile	.46	-.12	1.00	-.11
4 Educational Attainment	-.09	.05	-.11	1.00

Intercorrelations Between Measures for MOS 91A

Measure	Measure 1	2	3	4
1 HOPT	1.00	.01	.38	-.04
2 Time in Service	.01	1.00	-.16	.27
3 AFQT Percentile	.38	-.16	1.00	.03
4 Educational Attainment	-.04	.27	.03	1.00

Intercorrelations Between Measures for MOS 95B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.06	.30	-.11
2 Time in Service	.06	1.00	.00	.06
3 AFQT Percentile	.30	.00	1.00	-.06
4 Educational Attainment	-.11	.06	-.06	1.00

AFQT and Army job performance is based on our evaluation of five job performance factors: General Soldiering Proficiency, Core Technical Proficiency, Effort and Leadership, Personal Discipline, and Physical Fitness/Military Bearing. These factors were derived empirically from a complete set of soldier performance measures based on hands-on performance, written school and job knowledge and peer/supervisor performance ratings. The Army project scientists along with their scientific and policy advisors intended the hands-on test to be only one part of this integrated three-part strategy to measure representative job performance. The Army's project was designed such that a comprehensive task sampling procedure was used to define the population of tasks in each MOS, and then to select 30 job tasks to represent the population of the MOS tasks. The task lists were then reviewed by the proponent schools for completeness and representativeness of the occupation. Fifteen tasks requiring a high level of physical skill, a series of prescribed steps, and speed of performance were selected for hands-on testing. In other words, the appropriateness and feasibility of hands-on testing was a consideration in the selection of tasks for hands-on testing. Information about job performance was also obtained from written school and job knowledge tests as well as from a variety of peer and supervisory ratings. More detailed information is included in reports listed in the reference section of this chapter.

Sample Description. The Army's Project A sample has been described in detail in the Project's Annual Reports (see reference list) as well as earlier in this chapter.

Reliability Results. As reported in the Fifth Annual Report to Congress, the split-half reliability estimates for the hands-on, job knowledge, and school knowledge measures indicated a high degree of internal consistency in the measures for each MOS.

Similarly, the inter-rater reliability estimates for the Behaviorally Anchored Rating Scales (BARS) demonstrated moderate to strong reliabilities among raters. The fact that the reliability estimates for supervisor ratings are generally higher than the estimates for peer ratings may be attributed to the fact that supervisors have more experience rating individuals than non-supervisors.

Job Performance Measures and ASVAB Relationships. As presented in the last report to Congress, the intercorrelations among the criterion measures for each of the nine MOS are all non-zero and, for the most part, in the moderate range. This pattern of results is similar to that found in the field test data. These moderate correlations indicate that different aspects of the job performance domain are being measured by the different testing instruments.

Although the Army presented the results of its concurrent validation in the Fifth Annual Report to Congress, a summary is

supplied below. Briefly, statistical techniques were employed to determine criterion construct scores. Analyses resulted in the identification of five criterion constructs (factors):

1. Basic Soldiering Skills (use of basic weapons, first aid, etc.)

2. MOS Specific Technical Skills (document preparation for 71L; tank operation for 19E, etc.)

3. Exercise of Leadership, Effort and Self Development (the individual's willingness to perform the tasks and to be cooperative and supportive to other soldiers)

4. Maintaining Personal Discipline (adherence to Army regulations and traditions, commitment to high standards of personal conduct)

5. Military Bearing/Fitness (maintenance of appropriate military appearance and good physical condition)

Once the scores comprising each criterion factor (construct) were identified, the scores were weighted and summed within each factor to obtain a construct score. The correlations between the construct scores and AFQT scores for the nine MOS are presented in Table 6 and 7. The corrected correlations in Table 7 present "a more accurate picture of the validity of AFQT," as explained in Chapter 2 under "Validity Results." It should be noted that for 11B all skills are general soldiering skills. Examination of the correlations indicates that AFQT predicts General Soldiering Proficiency and Core Technical Proficiency

Table 6

Uncorrected Correlation Between AFQT and Criterion Construct Scores

MOS	N	General Soldiering Proficiency	Core Technical Proficiency	Effort and Leadership	Personal Discipline	Physical Fitness Military Bearing
11B	478	.42	---	.19	.15	-.03
13B	426	.34	.23	.11	.05	-.10
19E	367	.46	.26	.17	.13	-.06
31C	274	.39	.41	.08	.04	-.16
63B	461	.31	.24	.04	-.01	-.04
64C	486	.37	.23	.00	-.04	-.04
71L	407	.40	.43	.16	.10	.00
91A	378	.31	.34	.09	.07	-.01
95B	569	.30	.28	.13	.10	-.01

Table 7

Corrected Correlation Between AFQT and Criterion Construct Scores

MOS	N	General Soldiering Proficiency	Core Technical Proficiency	Effort and Leadership	Personal Discipline	Physical Fitness Military Bearing
11B	478	.58	---	.40	.25	.07
13B	426	.43	.33	.29	.13	.00
19E	367	.49	.23	.25	.08	-.02
31C	274	.59	.65	.10	.00	-.28
63B	461	.48	.51	.25	.02	-.10
64C	486	.60	.43	.10	-.14	-.10
71L	407	.57	.56	.28	.01	.05
91A	378	.68	.67	.21	.05	-.24
95B	569	.64	.61	.40	.31	-.12

Note: Numbers are corrected correlation coefficients. The correction for range restriction employed the multivariate correction procedure designated by the Joint Services Job Performance Measurement Working Group. (Working Group Minutes of the 9-10 July 1985 meeting, 26 August 1985.)

construct scores. The AFQT does not predict the other three construct scores; correlations between AFQT scores and Effort/Leadership, Personal Discipline and Fitness/Bearing scores are generally not statistically different from zero. These results are not unexpected since the AFQT measure was developed to predict success in training for first term military enlistment (selection) purposes.

As noted earlier, the Army's goals include the development and validation of new and/or improved selection and classification measures. Preliminary analyses of the Army project's additional (e.g., non-cognitive) predictor measures indicate that the criterion construct scores poorly predicted (Effort/Leadership, Personal Discipline and Physical Fitness/Military Bearing) by AFQT are well-predicted by other Project A measures.

SUMMARY AND CONCLUSIONS

The analyses presented in this chapter along with the data presented in the Fifth Annual report to Congress clearly support the position that ASVAB predicts the technical aspects of job

performance of first term soldiers, and, as such, can be used to link enlistment standards to job performance. In fact, the Army has already embarked upon research to examine various linkage issues and procedures. Nevertheless, perhaps one of the most striking conclusions that can be drawn from the Army's research on performance measurement to date is that jobs are indeed multi-dimensional. While ASVAB can predict the general soldiering and core technical (can-do) dimensions of performance, it is not a very good predictor of these will-do dimensions. It is clearly important to focus on all aspects of job performance to optimize the opportunity to improve significantly enlistment (selection and classification) systems.

REFERENCES

Campbell, J.P., and Harris, J.H. (1985, August). Criterion reduction and combination via a participative decision making panel. Paper presented at the Annual Convention of the American Psychological Association, Los Angeles, CA.

Campbell, R.C. (1985). Scorer training materials (RS Working Paper 85): Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR). (1984). Selecting job tasks for criterion tests of MOS proficiency. (RS-WP-84-25). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institute for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1983). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Research Report 1347). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1984). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Technical Report 660). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1985). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1985). Report on the Results of the Field Tests. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1986, under review). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics). (1983). A Report to the House Committee on Appropriations: Second Annual Report to Congress on Joint-Service Efforts to Link Standards for Enlistment to On-the-Job Performance.

Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics). (1984). A Report to the House Committee on Appropriations: Third Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standard to Job Performance.

Office of the Assistant Secretary of Defense (Force Management and Personnel). (1985). A Report to the House Committee on Appropriations: Fourth Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standards to Job Performance.

Office of the Assistant Secretary of Defense (Force Management and Personnel). (1986). A Report to the House Committee on Appropriations: Fifth Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standards to Job Performance.

Wigdor, A.K., and Green, B.F. (1986). Assessing the performance of enlisted personnel: Evaluation of a joint-service research project. Washington. D.C.: National Academy Press.

Wise, L.L., Wang, M., and Rossemissl, P.G. (1983). Development and validation of Army selection and classification measures Project A: Longitudinal Research Database Plan. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Service Research Efforts and Linkage Issues
Army Summary for DoD Chapter

Army

The Army is engaged in two research projects directed at linking enlistment standards to job performance. The first project, the Enlisted Personnel Allocation System (EPAS), begun in 1981 and scheduled for completion in 1989, takes what could be called a macro approach to the problem, as described below. The second project, also described below under Synthetic Validation began in March 1987 and is a three-year effort designed to examine the feasibility of applying synthetic validation and standard setting procedures within the Army's selection and classification system.

EPAS. The linking of soldier performance data to enlistment and assignment policy requires several kinds of information:

1. The applicant population

2. The jobs or tasks that need to be performed

3. Tradeoffs in the payoff from assigning applicants to different jobs

First of all, the information used in the linking of standards or quality requirements must be compatible across all alternatives. That is, the kinds of predictors should be generalizable across all Army applicants and jobs. There needs to be an assessment of the value of increasing performance in different jobs that can be used to make comparisons.

Second, some assessment is needed of the population available for selection and assignment. Policies need to be evaluated for an entire population against all job requirements, rather than against a small segment of the population or a limited number of jobs.

The Enlisted Personnel Allocation System (EPAS) provides a way of analyzing the impact of alternative policies and recruit populations on the Army. For example, different enlistment standards, high quality recruit supplies, and assignment payoffs can be simulated to find out what would happen to the Army's enlisted force with, for example:

1. different quality distributions

2. alternative quality pools

3. different enlistment or job standards

4. different predictor information

5. different job utility information

Synthetic Validation. In addition to its much-desribed work on Project A, and EPAS the Army is conducting a separate project to evaluate procedures for linking job performance to selection standards. The Army Synthetic Validation Project (SVP) is motivated by the need to establish selection standards for more than 250 entry-level enlisted MOS. Project A has developed measures of job performance for only 20 of these MOS.

The objective of the SVP is to evaluate the extent to which job performance prediction equations can be established synthetically, particularly in cases where empirical validation is not at all feasible. The synthetic validation approach involves examining ways in which performance in enlisted jobs can be decomposed into a set of common job components. In order to develop prediction equations synthetically, it must be possible to establish equations for predicting preformance on each component in the overall list of job components, and the equation for each component must generalize across all jobs for which the component is relevant. If these conditions hold, then overall prediction equations can be established as the weighted sum of prediction equations for individual job components.

A second part of the SVP concerns the development of procedures for setting job performance standards. If such standards can be reliably established, then selection standards can be based on the expected probability of achieving acceptable levels of job performance.

Selection and Classification Technical Area Working Paper RS-WP-88-9

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE: SEVENTH ANNUAL REPORT TO CONGRESS

Jane M. Arabian
Army Research Institute

Winnie Young and Lauress Wise
American Institutes for Research

September 1988

REVIEWED BY:

Clinton B. Walker

Clinton Walker
Research Psychologist

APPROVED BY:

Lawrence M Hanser

Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Cleared by:

for Newell K.Eaton, Director

Manpower and Personnel Research Laboratory

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

CHAPTER 3

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB PERFORMANCE

CURRENT STATUS AND ACCOMPLISHMENTS

The Army Research Institute is currently engaged in a large-scale, multi-year project to improve the Army selection and classification system and, thereby, increase the overall effectiveness of the force. The goal of the Army's program for increasing the efficiency of enlisted personnel selection and utilization is to enable the Army to meet its peacetime and mobilization missions through improved matching of individuals to Military Occupational Specialties (MOS). The research is aimed at developing comprehensive selection and classification procedures to validly predict performance in Army training and occupational specialties.

A complete description of the Army's research program and accomplishments has been published separately by the U.S. Army Research Institute for the Behavioral and Social Sciences in the annual reports for Project A, "Improving the Selection, Classification and Utilization of Army Enlisted Personnel" (1983, 1984, 1985, 1986). A detailed description of the project also appears in previous Annual Reports to Congress on Joint-Service Efforts to Link Standards of Enlistment to On-the-Job Performance (1983, 1984, 1985, 1986, 1987).

The Army's research focuses on 19 MOS. The MOS were selected to be representative of the Army and include all operational ASVAB aptitude area composites. Blacks, Whites, Hispanics, males, and females are present in these MOS in the same proportions as in total accessions. These MOS represent 44% of annual Army enlistments.

A number of measures, including measures of training success, service-wide job performance, and MOS-specific hands-on job performance, were developed for these MOS. For reasons of cost efficiency, not all measures were developed for all 19 MOS. All project criterion measures were developed for the following MOS and serve as the focus of this report:

1. 11B - Infantryman
2. 13B - Cannon Crewman
3. 19E - Tank Crewman
4. 31C - Radio Teletype Operator
5. 63B - Light Wheel Vehicle/Power Generation Mechanic
6. 64C - Motor Transport Operator
7. 71L - Administrative Specialist
8. 91A - Medical Specialist
9. 95B - Military Police

Measures of training success and service-wide performance were developed for the following specialties:

1. 12B - Combat Engineer
2. 16S - MANPADS Crewman
3. 27E - TOW/Dragon Repairman
4. 51B - Carpentry/Masonry Specialist
5. 54E - Chemical Operations Specialist
6. 55B - Ammunition Specialist
7. 67N - Utility Helicopter Repairman
8. 76W - Petroleum Supply Specialist
9. 76Y - Supply Specialist
10. 94B - Food Service Specialist

A complete report of accomplishments and current status of the Army's research to link enlistment standards to job performance has been published by the Army Research Institute (1983, 1984, 1985, 1986) and is available for distribution. What follows includes a summary of the Army's accomplishments in the area of job performance measurement.

As reported last year in the Annual Report to Congress the Army has completed the concurrent validation phase, i.e., joint administration of predictor and criterion measures, of its project. The research not only established the feasibility of measuring job performance but also identified five performance constructs, described later in this chapter, thereby providing support for the notion that job performance is multi-dimensional. In addition, the research established that there is a positive relationship between ASVAB and job performance during the first term of service. More specifically, the higher an applicant's ASVAB score the better he or she performs the technical or "can-do" aspects of the job. The correlations ranged from .26 to .70. The "will-do" or motivational factors are not well-predicted by ASVAB, but the Army's new predictor measures are expected to improve the prediction of the non-cognitive job performance dimensions.

The Army is currently engaged in the longitudinal validation phase of its project. Fifty-three thousand Army entrants were administered the Army project's predictor battery at eight reception battalions. Collection of training (school) knowledge measures, administered to soldiers in the sample at the end of Advanced Individual Training (AIT), ended in May of 1988. Administration of the job performance measures is now being conducted.

Implementation of the project's products currently includes three efforts. First, with respect to the use of ASVAB, modifications have been made to the aptitute area composites to improve can-do assignment efficiency. Forty-four MOS are in the process of being reassigned to different entry composites in order to: increase validity, increase consistency of composite use within a career management field, and increase the accuracy

of prediction for women and blacks. The second major implementation effort involves one of the new predictors, Assessment of Background and Life Experiences (ABLE), a temperament/personal history instrument. The Army is presently planning for an Initial Operational Test and Evaluation (IOT&E) of ABLE. ABLE will help the Army to predict better the job dimensions of: Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing. The third implementation is designed to better capitalize on individual psychomotor and spatial attributes in MOS which require them for effective gunnery performance.

RESEARCH DESIGN

Types of Measures Developed

Hands-on performance measures, job knowledge tests, and performance rating scales were developed for training success, service-wide performance, and MOS-specific performance for the MOS listed previously. The Army's rationale for the development of multiple measures of job performance is based upon the knowledge that a soldier's job is multi-faceted and there are different aspects of job performance. Therefore, the Army's research project has developed different kinds of tests to assess these different aspects of job performance and thereby obtain information about the domain of job performance behaviors. A more complete description of these measures appears in the Project A Annual Reports (1983, 1984, 1985, 1986) and in previous Annual Reports to Congress (1983, 1984, 1985, 1986).

Data Collection Procedures

Data collection began with the briefing of local military commanders, examination of the test sites, equipment and supplies, training of test administrators and scorers, and a dry run of testing procedures. Test site managers were appointed to supervise the actual data collections and were responsible for controlling the quality and flow of data from the testing.

The concurrent validation data collection began in June and ended in November 1985. As reported in the Fourth Annual Report to Congress, data were collected at fourteen different sites. The data from 5200 soldiers in 9 MOS, who were tested on all the project criterion measures, were entered into the research data base. Data from an additional 4000 soldiers in the 10 other MOS, who were tested only with the measures of training success and service-wide performance were also entered. Analyses began as soon as all the data were collected and entered in the longitudinal research data base.

RESULTS OF DATA ANALYSES

Common Data Analyses

Sample Description. The analyses are based on data collected during the Army's concurrent validation data collection in 1985. Soldiers were in Skill Level One, entry-level positions in their respective MOS primarily with 18-24 months experience in the Army.

Table 1 presents descriptive statistics for the variables included in these analyses. The hands-on performance test (HOPT) scores have been converted to standard T-scores in order to allow comparison across MOS. Time in service is the number of months, averaged across soldiers, of Army service. In accordance with the Army's sampling plan, the means are between 18 and 24 months, with small variances (standard deviations). Average AFQT percentile score for each MOS sample is also presented, along with the number of soldiers classified as high school diploma graduates (HSDG) or non-high school diploma graduates (NHSDG).

Table 2 also presents descriptive statistics for the sample, this time broken into four groups for time in service and two AFQT groups, upper 50 percentiles (I-IIIA) and lower 50 percentiles (IIIB-IV). Again, given the design of the Army's project, few soldiers were expected in the 1-12, 25-36, and 37+ Time-in-Service groups. Therefore, cells with fewer than 25 observations have been marked "N/A".

Reliability Results. Two indices of reliability, split-half and coefficient alpha, for HOPT were computed and are presented in Table 3. Both uncorrected and corrected reliability estimates are included along with the standard error of measurement. The table also includes the number of tasks in the HOPT scores for these common data analyses.

Hands-On Performance Test, Aptitude, Education, and Job Experience Relationships. Intercorrelations among the measures are presented for each MOS. Uncorrected correlations are depicted in Table 4. The correlations in Table 5 and elsewhere, as noted in this chapter, have been corrected for range restriction using a multivariate correction procedure as agreed upon by the Joint Service Job Performance Measurement Working Group for the purposes of these analyses.

Service-specific Data Analyses

Simply for purposes of comparison, Table 6 presents the correlations between the hands-on performance test scores and the Army's ASVAB classification (aptitude area or AA) composite for each of the nine MOS. Since the AA composites are used to classify individuals for entry into specific occupations, it is

Table 1

Descriptive Statistics for HOPT, Time in Service, Aptitude, and Educational Attainment

MOS		HOPT (Standard T-Score)	Time in Service (Months)	AFQT (Percentile)	Educational Attainment (N)	
	Mean	50	20.4	55.9	423	HSDG
11B	SD	10	5.6	21.8	55	NHSDG
	N	491	478	491		
	Mean	50	21.8	46.8	353	HSDG
13B	SD	10	6.1	20.2	73	NHSDG
	N	434	425	464		
	Mean	50	19.4	54.7	317	HSDG
19E	SD	10	4.6	22.0	50	NHSDG
	N	378	367	394		
	Mean	50	20.5	56.9	245	HSDG
31C	SD	10	5.6	20.9	29	NHSDG
	N	275	274	289		
	Mean	50	19.9	49.6	397	HSDG
63B	SD	10	5.0	19.6	64	NHSDG
	N	443	461	478		
	Mean	50	20.0	42.8	420	HSDG
64C	SD	10	4.6	19.5	66	NHSDG
	N	481	486	507		
	Mean	50	22.3	54.1	401	HSDG
71L	SD	10	4.5	20.3	7	NHSDG
	N	416	408	427		
	Mean	50	20.6	61.2	310	HSDG
91A	SD	10	5.4	16.9	68	NHSDG
	N	392	378	392		
	Mean	50	21.3	63.4	525	HSDG
95B	SD	10	5.0	15.1	44	NHSDG
	N	577	569	597		

Table 2

Mean HOPT Scores (Standardized T-Scores) by Aptitude and Job Experience Levels for 9 Jobs

Time in Service (Months)		MOS 11B AFQT I-IIIA	MOS 11B AFQT IIIB-IV	MOS 13B AFQT I-IIIA	MOS 13B AFQT IIIB-IV	MOS 19E AFQT I-IIIA	MOS 19E AFQT IIIB-IV
1-12	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	7	6	1	5	7	4
13-24	Mean	51.8	48.1	49.9	47.9	51.0	46.3
	SD	10.3	9.6	10.1	9.5	9.0	11.1
	N	214	154	115	151	163	121
25-36	Mean	50.8	48.3	53.4	52.3	55.7	51.1
	SD	9.4	9.3	11.4	8.9	7.1	9.0
	N	42	46	46	66	32	25
37+	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	6	3	5	6	0	1
Total Group	Mean	51.5	47.9	50.8	49.4	52.2	46.9
	SD	10.0	9.7	10.5	9.6	9.0	10.9
	N	269	209	167	228	202	151

Time in Service (Months)		MOS 31C AFQT I-IIIA	MOS 31C AFQT IIIB-IV	MOS 63B AFQT I-IIIA	MOS 63B AFQT IIIB-IV	MOS 64C AFQT I-IIIA	MOS 64C AFQT IIIB-IV
1-12	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	10	5	6	7	4	7
13-24	Mean	51.9	47.6	51.3	48.3	51.7	48.4
	SD	10.0	9.1	9.9	9.9	9.6	9.9
	N	136	78	172	177	106	248
25-36	Mean	N/A	N/A	N/A	51.9	55.1	49.7
	SD	---	---	---	10.3	8.6	10.8
	N	17	9	16	46	36	60
37+	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	2	3	1	3	0	0
Total Group	Mean	51.6	47.0	51.1	49.1	52.7	48.7
	SD	10.3	8.9	10.1	9.9	9.5	10.0
	N	165	95	195	233	146	315

Table 2 (Continued)

Mean HOPT Scores (Standardized T-Scores) by Aptitude and Job Experience Levels for 9 Jobs

Time in Service (Months)		MOS 71L AFQT I-IIIA	MOS 71L AFQT IIIB-IV	MOS 91A AFQT I-IIIA	MOS 91A AFQT IIIB-IV	MOS 95B AFQT I-IIIA	MOS 95B AFQT IIIB-IV
1-12	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	7	3	10	4	19	5
13-24	Mean	52.9	47.2	51.1	47.1	50.1	48.1
	SD	9.7	9.8	9.7	10.3	10.1	9.7
	N	173	125	209	74	317	84
25-36	Mean	52.4	47.1	50.4	N/A	52.0	49.3
	SD	8.5	8.7	10.5	---	10.5	8.7
	N	45	42	46	22	99	25
37+	Mean	N/A	N/A	N/A	N/A	N/A	N/A
	SD	---	---	---	---	---	---
	N	2	1	2	1	1	0
Total Group	Mean	52.8	47.1	51.1	46.9	50.5	48.3
	SD	9.5	9.6	9.8	9.9	10.1	9.4
	N	227	171	277	101	436	114

Table 3

Reliability Estimates and Standard Errors of Measurement for Hands-on Performance Tests

MOS	Number of Tasks		Split-Half	Coefficient Alpha	SEM[a]
11B	14	ucoef.[b]	.56	.55	6.71
		ccoef.[c]	.57	.57	
13B	17	ucoef.	.76	.70	5.52
		ccoef.	.76	.70	
19E	15	ucoef.	.57	.56	6.66
		ccoef.	.59	.57	
31C	15	ucoef.	.81	.69	5.58
		ccoef.	.83	.72	
63B	15	ucoef.	.47	.40	7.74
		ccoef.	.47	.40	
64C	16	ucoef.	.66	.60	6.32
		ccoef.	.68	.63	
71L	15	ucoef.	.73	.68	5.65
		ccoef.	.76	.72	
91A	17	ucoef.	.59	.72	5.31
		ccoef.	.62	.74	
95B	16	ucoef.	.63	.52	6.90
		ccoef.	.65	.55	

[a]Standard Error of Measurement calculated with the Coefficient Alpha reliability estimates and HOPT standard T-Score standard deviations.

[b]Uncorrected reliability estimate.

[c]Corrected reliability estimate.

Table 4
Uncorrected Intercorrelations

Intercorrelations Between Measures for MOS 11B

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.06	.25	.12
2 Time in Service	-.06	1.00	-.01	.05
3 AFQT Percentile	.25	-.01	1.00	.08
4 Educational Attainment[a]	.12	.05	.08	1.00

Intercorrelations Between Measures for MOS 13B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.18	.13	.02
2 Time in Service	.18	1.00	-.01	.04
3 AFQT Percentile	.13	-.01	1.00	-.11
4 Educational Attainment	.02	.04	-.11	1.00

Intercorrelations Between Measures for MOS 19E

Measure	Measure 1	2	3	4
1 HOPT	1.00	.26	.26	-.08
2 Time in Service	.26	1.00	-.02	-.08
3 AFQT Percentile	.26	-.02	1.00	.01
4 Educational Attainment	-.08	-.08	.01	1.00

Intercorrelations Between Measures for MOS 31C

Measure	Measure 1	2	3	4
1 HOPT	1.00	.01	.34	.04
2 Time in Service	.01	1.00	-.02	-.03
3 AFQT Percentile	.34	-.02	1.00	-.02
4 Educational Attainment	.04	-.03	-.02	1.00

Intercorrelations Between Measures for MOS 63B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.07	.13	-.05
2 Time in Service	.07	1.00	-.11	.12
3 AFQT Percentile	.13	-.11	1.00	-.08
4 Educational Attainment	-.05	.12	-.08	1.00

Table 4 (Continued)
Uncorrected Intercorrelations

Intercorrelations Between Measures for MOS 64C

Measure	Measure 1	2	3	4
1 HOPT	1.00	.04	.24	.06
2 Time in Service	.04	1.00	.07	-.11
3 AFQT Percentile	.24	.07	1.00	-.10
4 Educational Attainment	.06	-.11	-.10	1.00

Intercorrelations Between Measures for MOS 71L

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.02	.35	.06
2 Time in Service	-.02	1.00	-.10	-.05
3 AFQT Percentile	.35	-.10	1.00	.05
4 Educational Attainment	.06	-.05	.05	1.00

Intercorrelations Between Measures for MOS 91A

Measure	Measure 1	2	3	4
1 HOPT	1.00	.04	.28	.05
2 Time in Service	.04	1.00	-.08	-.28
3 AFQT Percentile	.28	-.08	1.00	.04
4 Educational Attainment	.05	-.28	.04	1.00

Intercorrelations Between Measures for MOS 95B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.06	.23	.11
2 Time in Service	.06	1.00	-.01	-.06
3 AFQT Percentile	.23	-.01	1.00	.09
4 Educational Attainment	.11	-.06	.09	1.00

Note. [a]For purposes of these analyses the coding for Educational Attainment was: 1 = Non-high school degree; 2 = High school degree.

Table 5
Intercorrelations Corrected for Range Restriction

Intercorrelations Between Measures for MOS 11B

Measure	1	2	3	4
1 HOPT	1.00	-.08	.34	.14
2 Time in Service	-.08	1.00	-.04	.04
3 AFQT Percentile	.34	-.04	1.00	.13
4 Educational Attainment[a]	-.14	.04	.13	1.00

Intercorrelations Between Measures for MOS 13B

Measure	1	2	3	4
1 HOPT	1.00	.18	.15	.02
2 Time in Service	.18	1.00	.01	.04
3 AFQT Percentile	.15	.01	1.00	-.09
4 Educational Attainment	.02	.04	-.09	1.00

Intercorrelations Between Measures for MOS 19E

Measure	1	2	3	4
1 HOPT	1.00	.25	.31	-.06
2 Time in Service	.25	1.00	.00	-.08
3 AFQT Percentile	.31	.00	1.00	.02
4 Educational Attainment	-.06	-.08	.02	1.00

Intercorrelations Between Measures for MOS 31C

Measure	1	2	3	4
1 HOPT	1.00	.02	.51	.06
2 Time in Service	.02	1.00	.01	-.03
3 AFQT Percentile	.51	.01	1.00	.04
4 Educational Attainment	.06	-.03	.04	1.00

Intercorrelations Between Measures for MOS 63B

Measure	1	2	3	4
1 HOPT	1.00	.05	.13	-.07
2 Time in Service	.05	1.00	-.14	.13
3 AFQT Percentile	.13	-.14	1.00	.13
4 Educational Attainment	-.07	.13	-.08	1.00

Table 5 (Continued)
Intercorrelations Corrected for Range Restriction

Intercorrelations Between Measures for MOS 64C

Measure	Measure 1	2	3	4
1 HOPT	1.00	.09	.39	.01
2 Time in Service	.09	1.00	.12	-.13
3 AFQT Percentile	.39	.12	1.00	-.11
4 Educational Attainment	.01	-.13	-.11	1.00

Intercorrelations Between Measures for MOS 71L

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.09	.49	.11
2 Time in Service	-.09	1.00	-.19	-.07
3 AFQT Percentile	.49	-.19	1.00	.12
4 Educational Attainment	.11	-.07	.12	1.00

Intercorrelations Between Measures for MOS 91A

Measure	Measure 1	2	3	4
1 HOPT	1.00	-.01	.46	.03
2 Time in Service	-.01	1.00	-.15	-.28
3 AFQT Percentile	.46	-.15	1.00	.00
4 Educational Attainment	.03	-.28	.00	1.00

Intercorrelations Between Measures for MOS 95B

Measure	Measure 1	2	3	4
1 HOPT	1.00	.07	.49	.13
2 Time in Service	.07	1.00	.01	-.06
3 AFQT Percentile	.49	.01	1.00	.12
4 Educational Attainment	.13	-.06	.12	1.00

Note. [a]For purposes of these analyses the coding for Educational Attainment was: 1 = Non-high school degree; 2 = High school degree.

not surprising that these correlations are somewhat higher than the correlations between the hands-on scores and AFQT.

Table 6

Correlations of AA Classification Composites and HOPT

MOS	AA Composite	N	Uncorrected Correlation	Corrected Correlation
11B	CO	478	.31	.41
13B	FA	395	.18	.20
19E	CO	353	.29	.37
31C	SC	260	.36	.53
63B	MM	428	.24	.28
64C	OF	461	.37	.52
71L	CL	398	.37	.50
91A	ST	378	.35	.57
95B	ST	550	.32	.57

We believe, however, that the most appropriate data showing the link between ASVAB and Army job performance is based on our evaluation of five job performance factors: General Soldiering Proficiency, Core Technical Proficiency, Effort and Leadership, Personal Discipline, and Physical Fitness/Military Bearing. These factors were derived empirically from a complete set of soldier performance measures based on hands-on performance, written school and job knowledge and peer/supervisor performance ratings. The Army project scientists along with their scientific and policy advisors intended the hands-on test to be only one part of this integrated three-part strategy to measure representative job performance. The Army's project was designed such that a comprehensive task sampling procedure was used to define the population of tasks in each MOS, and then to select 30 job tasks to represent the population of the MOS tasks. The task lists were then reviewed by the proponent schools for completeness and representativeness of the occupation. Fifteen tasks requiring a high level of physical skill, a series of prescribed steps, and speed of performance were selected for hands-on testing. In other words, the appropriateness and feasibility of hands-on testing was a consideration in the selection of tasks for hands-on testing. Information about job performance was also obtained from written school and job knowledge tests as well as from a variety of peer and supervisory ratings. More detailed information is included in reports listed in the reference section of this chapter.

Sample Description. The Army's Project A sample has been

described in detail in the Project's Annual Reports (see reference list) as well as earlier in this chapter.

Reliability Results. As reported in the Fifth Annual Report to Congress, the split-half reliability estimates for the hands-on, job knowledge, and school knowledge measures indicated a high degree of internal consistency in the measures for each MOS.

Similarly, the inter-rater reliability estimates for the Behaviorally Anchored Rating Scales (BARS) demonstrated moderate to strong reliabilities among raters. The fact that the reliability estimates for supervisor ratings are generally higher than the estimates for peer ratings may be attributed to the fact that supervisors have more experience rating individuals than non-supervisors.

Job Performance Measures and ASVAB Relationships. As presented in the last two reports to Congress, the intercorrelations among the criterion measures for each of the nine MOS are all non-zero and, for the most part, in the moderate range. This pattern of results is similar to that found in the field test data. These moderate correlations indicate that different aspects of the job performance domain are being measured by the different testing instruments.

Although the Army presented the results of its concurrent validation in the Fifth Annual Report to Congress, a summary is supplied below. Briefly, statistical techniques were employed to determine criterion construct scores. Analyses resulted in the identification of five criterion constructs (factors):

1. Basic Soldiering Skills (use of basic weapons, first aid, etc.)

2. MOS Specific Technical Skills (document preparation for 71L; tank operation for 19E, etc.)

3. Exercise of Leadership, Effort and Self Development (the individual's willingness to perform the tasks and to be cooperative and supportive to other soldiers)

4. Maintaining Personal Discipline (adherence to Army regulations and traditions, commitment to high standards of personal conduct)

5. Military Bearing/Fitness (maintenance of appropriate military appearance and good physical condition)

Once the scores comprising each criterion factor (construct) were identified, the scores were weighted and summed within each factor to obtain a construct score. The correlations between the construct scores and AFQT scores for the nine MOS are presented in Table 7 and 8. It should be noted that for 11B all skills are general soldiering skills. Examination of the correlations

indicates that AFQT predicts General Soldiering Proficiency and Core Technical Proficiency construct scores. The AFQT does not

Table 7

Uncorrected Correlation Between AFQT and Criterion Construct Scores

MOS	N	General Soldiering Proficiency	Core Technical Proficiency	Effort and Leadership	Personal Discipline	Physical Fitness Military Bearing
11B	478	.46	---	.19	.16	-.06
13B	395	.35	.21	.09	.03	-.11
19E	353	.48	.40	.20	.15	-.05
31C	260	.45	.41	.12	.05	-.14
63B	428	.35	.31	.04	.04	-.05
64C	461	.42	.28	.05	.00	-.06
71L	398	.42	.45	.17	.13	.01
91A	378	.36	.38	.07	.09	-.07
95B	550	.34	.29	.07	.07	-.07

Table 8

Corrected Correlation Between AFQT and Criterion Construct Scores

MOS	N	General Soldiering Proficiency	Core Technical Proficiency	Effort and Leadership	Personal Discipline	Physical Fitness Military Bearing
11B	478	.57	---	.28	.24	-.05
13B	395	.43	.26	.20	.12	.02
19E	353	.56	.45	.29	.20	-.02
31C	260	.56	.60	.16	.05	-.21
63B	428	.45	.39	.14	.10	-.01
64C	461	.58	.39	.21	.06	.00
71L	398	.51	.56	.21	.09	.07
91A	378	.58	.60	.16	.16	-.10
95B	550	.64	.57	.26	.22	-.12

predict the other three construct scores; correlations between AFQT scores and Effort/Leadership, Personal Discipline and Fitness/Bearing scores are generally not statistically different from zero. These results are not unexpected since the AFQT measure was developed to predict success in training for first term military enlistment (selection) purposes.

Similarly, Tables 9 and 10 present the correlations between the AA Composite Scores and the Army's Criterion Construct Scores. Again, the ASVAB AA scores predict the two proficiency constructs better than the remaining three constructs.

Table 9

Uncorrected Correlation Between AA Classification Composite and Criterion Construct Scores

MOS	N	AA	General Soldiering Proficiency	Core Technical Proficiency	Effort & Leader-ship	Personal Discipline	Physical Fitness/ Military Bearing
11B	478	CO	.48	--	.23	.20	-.03
13B	395	FA	.35	.24	.16	.09	-.08
19E	353	CO	.47	.42	.26	.13	.01
31C	260	SC	.46	.40	.19	.10	.00
63B	428	MM	.33	.51	.20	.07	-.13
64C	461	OF	.49	.36	.18	.02	-.06
71L	398	CL	.43	.46	.18	.14	.00
91A	378	ST	.45	.44	.10	.08	-.13
95B	550	ST	.41	.29	.09	.09	-.09

Table 10

Corrected Correlation Between AA Classification Composite and Criterion Construct Scores

MOS	N	AA	General Soldiering Proficiency	Core Technical Proficiency	Effort & Leader-ship	Personal Discipline	Physical Fitness/ Military Bearing
11B	478	CO	.60	--	.32	.27	-.03
13B	395	FA	.44	.29	.25	.16	.03
19E	353	CO	.57	.51	.34	.16	.02
31C	260	SC	.58	.59	.23	.10	-.23
63B	428	MM	.44	.62	.29	.13	-.13
64C	461	OF	.65	.49	.33	.07	-.07
71L	398	CL	.52	.57	.22	.10	.08
91A	378	ST	.64	.64	.19	.15	-.15
95B	550	ST	.70	.57	.28	.23	-.14

As noted earlier, the Army's goals include the development and validation of new and/or improved selection and classification measures. Analyses of the Army project's additional (e.g., non-cognitive) predictor measures indicate that the criterion

construct scores poorly predicted (Effort/Leadership, Personal Discipline and Physical Fitness/Military Bearing) by ASVAB are well-predicted by other Project A measures.

SUMMARY AND CONCLUSIONS

The analyses presented in this chapter along with the data presented in the Fifth and Sixth Annual reports to Congress clearly support the position that ASVAB predicts the technical aspects of job performance of first term soldiers, and, as such, can be used to link enlistment standards to job performance. In fact, the Army has already embarked upon research to examine various linkage issues and procedures. Nevertheless, perhaps one of the most striking conclusions that can be drawn from the Army's research on performance measurement to date is that jobs are indeed multi-dimensional. While ASVAB can predict the general soldiering and core technical (can-do) dimensions of performance, it is not a very good predictor of the will-do dimensions. It is clearly important to focus on all aspects of job performance to optimize the opportunity to improve significantly enlistment (selection and classification) systems.

Emerging Applications

In addition to the research presented in this report, the Army is also studying the relationship between enlistment standards and unit readiness. Unit performance is being examined in these specialties: Field Artillery, Armor, Infantry, Air Defense "Maintenance", and Signal. Preliminary results indicate positive relationships between unit performance and enlistment standards, particularly AFQT scores, consistent with the findings from Project A. These research efforts are discussed in the Army Appendix to this Chapter.

REFERENCES

Campbell, J.P., and Harris, J.H. (1985, August). Criterion reduction and combination via a participative decision making panel. Paper presented at the Annual Convention of the American Psychological Association, Los Angeles, CA.

Campbell, R.C. (1985). Scorer training materials (RS Working Paper 85): Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR). (1984). Selecting job tasks for criterion tests of MOS proficiency. (RS-WP-84-25). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institute for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1983). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Research Report 1347). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1984). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Technical Report 660). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1985). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1985). Report on the Results of the Field Tests. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), Personnel Decisions Research Institute (PDRI), and Army Research Institute (ARI). (1986). Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report. (ARI Technical Report 813101). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics). (1983). A Report to the House Committee on Appropriations: Second Annual Report to Congress on Joint-Service Efforts to Link Standards for Enlistment to On-the-Job Performance.

Office of the Assistant Secretary of Defense (Manpower, Installations, and Logistics). (1984). A Report to the House Committee on Appropriations: Third Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standard to Job Performance.

Office of the Assistant Secretary of Defense (Force Management and Personnel). (1985). A Report to the House Committee on Appropriations: Fourth Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standards to Job Performance.

Office of the Assistant Secretary of Defense (Force Management and Personnel). (1986). A Report to the House Committee on Appropriations: Fifth Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standards to Job Performance.

Office of the Assistant Secretary of Defense (Force Management and Personnel). (1987). A Report to the House Committee on Appropriations: Sixth Annual Report to Congress on Joint-Service Efforts to Link Enlistment Standards to Job Performance.

Wigdor, A.K., and Green, B.F. (1986). Assessing the performance of enlisted personnel: Evaluation of a joint-service research project. Washington. D.C.: National Academy Press.

Wise, L.L., Wang, M., and Rossemissl, P.G. (1983). Development and validation of Army selection and classification measures Project A: Longitudinal Research Database Plan. (ARI Technical Report 1346). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-22

The Incremental Validity of the Figural Reasoning and Mazes Tests in Project A's 1985 Concurrent Validation

Henry H. Busciglio
Army Research Institute

November 1989

REVIEWED BY:

Frances C. Grafton
Statistician
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

INTRODUCTION

The Figural Reasoning and Mazes tests are two of the cognitive paper-and-pencil measures developed in the Army's Project A to assess abilities of potential soldiers not covered by the general cognitive domain of the Armed Services Vocational Aptitude Battery (ASVAB). Factor analyses conducted during the construction of Project A predictor composites (cf. Campbell, 1988) found that both instruments loaded on a common factor, termed "spatial abilities"; however, these tests appear to tap somewhat different domains of abilities. The figural reasoning test appears to measure general intelligence, in terms of inductive reasoning and working memory. In contrast, the mazes test has been classified by some (but not all) experts as a test of spatial visualization.

A larger investigation (Busciglio, in preparation) assessed the utility of the Project A tests, including Figural Reasoning and Mazes, for incrementing the validity of ASVAB across a variety of criterion measures in nine high-fill Military Occupational Specialities (MOS), or Army jobs. Project A criteria used were the total score on written tests (summed across the School and Job Knowledge tests) and hands-on tests (summed across all hands-on tasks), as well as scores on selected individual tasks in the written (e.g., Target Identification) and hands-on (e.g., Navigation, Operating Howitzer Sights, and Tank Gunnery) modes. Also included were two performance constructs from Project A which divided the written and hands-on tasks into two composite scores: General Soldiering Proficiency and Core Technical (i.e., MOS-Specific) Proficiency.

In addition to these outcome measures from Project A, scores from written tests developed annually by the Army were used. These instruments, Skill Qualification Tests (SQT), are MOS-specific measures administered to incumbents periodically to certify their continuing job proficiency. Although the SQT covers the same skill domains as the Project A criteria, sampling of the tasks in these domains differs in unsystematic ways from that used in Project A.

METHOD

In the larger effort, a series of stepwise multiple regression analyses were conducted separately for each criterion measure and MOS for which hands-on testing was done. The SPSS-X regression procedure that was employed made it possible to control for restriction-of-range in the ASVAB subtests used by the Army for selection and classification, as well as any implicit restriction in the range of predictors or criteria which are correlated with these ASVAB subtests. This was accomplished by using as input a matrix of estimated population (i.e., unrestricted) covariances among all predictors and criteria,

employing the Lawley correction formula presented in Lord and Novick (1968, pp. 144-148).

The stepwise analyses first determined an _optimal_ prediction equation for ASVAB scores by entering all nine subtests as a block, then using a backward elimination technique to remove subtests that were not individually significant (at the .05 level). Then, blocks of either spatial or perceptual-psychomotor tests were added and the backward elimination procedure determined which of these added _individually_ to the R^2 for the ASVAB scores. To ensure that these predictors showed true incremental validity, the ASVAB scores found to be significant in the initial equations were not subject to removal when the Project A tests were added. In the final stage of analysis, spatial tests were added to the equations containing significant ASVAB and perceptual-psychomotor scores, and perceptual-psychomotor tests were added to equations composed of significant ASVAB and spatial scores. Once again, the added tests were retained only if individually significant, while the predictors already in the equation were not subject to backward elimination.

RESULTS

Using the above procedures, the incremental utility of individual Project A tests was assessed separately for each MOS. Table 1 shows results for the Figural Reasoning and Mazes tests, in terms of individual semi-partial correlations for the significant predictors. For each criterion and MOS, predictors can be significant in _two_ equations: at the second stage of analysis when added, with the other spatial tests, to the equations containing only significant ASVAB subtests, and at the third stage, when added to significant ASVAB and perceptual-psychomotor scores.

As Table 1 shows, the Mazes test was found to be a poor incremental predictor, across all criteria and MOS. In contrast, the Figural Reasoning test appeared to be highly useful for improving the prediction of quite a few criterion measures, particularly Total Score on School and Job Knowledge Tests, Total Score on Hands-On Tests, General Soldiering Proficiency, Core Technical Proficiency, and Determining Grid Coordinates.

The positive results for the Figural Reasoning test must be viewed in light of its possible redundancy with another strong spatial predictor, Assembling Objects. For example, Peterson (1987) found that the two tests showed the highest intercorrelation among the six spatial ability tests and loaded highly on the same factor when spatial tests were factor analyzed (see pp. 81, 88-89 from the Army's report to the Technical Advisory Service Panel). The redundancy between the two tests was assessed by examining the Busciglio (in preparation) results

to determine the number of MOS for which <u>both</u> tests were significant incremental predictors of the criterion measures. It was found, as shown in Table 2, that in the 49 cases where at least one of the two tests was a significant predictor, there were 17 instances where <u>both</u> measures entered the prediction equation.

DISCUSSION

To properly interpret the results above, it is important to note that there is a certain degree of redundancy among some criteria. First of all, written and hands-on tests contributed variance to either the Total Score on School and Job Knowledge Tests or the Total Score on Hands-On Tests, as well as to General Soldiering and Core Technical Proficiency. Along the same lines, some of the more specific criterion measures (e.g., Navigation, Target Identification) form a small part of the overall success scores, such as General Soldiering and Core Technical Proficiency. Another methodological consideration is that the Figural Reasoning and Mazes tests are in a paper-and-pencil format, introducing the question of "method variance" into their correlations with the paper-and-pencil criterion measures (although variance due to <u>verbal</u> ability is perhaps only a small portion of total variance on the Figural Reasoning and Mazes test).

The cautions above notwithstanding, the results of the analyses indicate a reasonably high probability that in the future the Figural Reasoning test will be shown to be more useful than the Mazes test for incrementing the validity of ASVAB for predicting performance criteria across a wide variety of MOS. These results are directly relevant to the selection of tests to be used in the inter-Service Enhanced Computer Assisted Testing (ECAT) research.

REFERENCES

Busciglio, H.H. (in preparation). The incremental validity of spatial and perceptual-psychomotor tests in Project A's 1985 concurrent validation (ARI Technical Report in preparation). Alexandria, VA: U.S. Army Research Institute.

Campbell, J.P. (Ed.). (1988). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). Alexandria, VA: U.S. Army Research Institute.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley Publishing Co.

Peterson, N.G. (Ed.). (1987). Development and field test of the Trial Battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute.

Table 1

Significant Semi-partial Correlations for Figural Reasoning and Mazes Tests as Incremental Predictors of Criteria

	MOS								
	11B[a]	13B	19E	31C	63B	64C	71L	91A	95B
Total Score on School and Job Knowledge Tests:									
Figural Reas.	.06		.07	.07		.08	.07	.06	.06
	.06			.07		.07	.07		.06
Mazes					.05				
					.05				
Total Score on Hands-On Tests:									
Figural Reas.						.08	.12	.13	.09
							.14	.12	.07
Mazes		.07							
General Soldiering Proficiency:									
Figural Reas.	na		.10	.14		.07		.08	.07
	na		.12	.14					.07
Mazes					(none)				
Core Technical Proficiency:									
Figural Reas.	.07					.07	.09	.13	
	.07						.09	.13	
Mazes							-.05		
							-.08		
Skill Qualification Test Score:									
Figural Reas.	.08				.08		.08	na	
	.10				.07		.10	na	
Mazes					(none)				
Navigation:									
Figural Reas.	na		.14				.09		
	na		.14				.13		
Mazes	na								.08
	na								
Target Identification (excludes MOS 63B and 71L):									
Figural Reas.					(none)				
Mazes					(none)				

Table 1 (cont.)

Significant Semi-partial Correlations for Figural Reasoning and Mazes Tests as Incremental Predictors of Criteria

	MOS								
	11B[a]	13B	19E	31C	63B	64C	71L	91A	95B
Operating Howitzer Sights (MOS 13B only):									
Figural Reas.		(none)							
Mazes		(none)							
Tank Gunnery (MOS 19E only):									
Figural Reas.			(none)						
Mazes			(none)						
Determining Grid Coordinates:									
Figural Reas.	na	na	.14	.08	na	.09	.10	.07	
	na	na	.14		na	.09	.10		
Mazes					(none)				

Note. Semipartial r's for significant predictors only, separately for two orders of entry (ASVAB-Spatial-Psychomotor and ASVAB-Psychomotor-Spatial); "na" is shown if a criterion was not tested in the MOS.

[a] The MOS were as follows:
- 11B - Infantrymen (N=491)
- 13B - Cannon Crewmen (N=464)
- 19E - Armor Crewmen (N=394)
- 31C - Single Channel Radio Operator (N=289)
- 63B - Light Wheel Vehicle Mechanic (N=478)
- 64C - Motor Transport Operator (N=507)
- 71L - Administrative Specialist (N=427)
- 91A - Medical Specialist (N=392)
- 95B - Military Police (N=597)

Table 2

Number of MOS for which Assembling Objects and Figural Reasoning are Significant[a] Predictors of Criteria

	Both	Assembling Objects	Figural Reasoning	Neither	Total
School/Job Know Tests:	6	2	1	0	9
Hands-On Tests:	1	1	3	4	9
Gen Soldiering Prof:	4	3	1	0	8
Core Technical Prof:	2	3	2	2	9
Skill Qual Test:	1	4	2	1	8
Navigation:	1	3	1	3	8
Target Identification:	0	1	0	6	7
Align Howitzer Sights:	0	1	0	0	1
Tank Gunnery:	0	0	0	1	1
Deter Grid Coordinates:	2	1	3	0	6
TOTAL	17	19	13	17	66

[a] In either or both orders of entry.

Working Paper

THE USEFULNESS OF PROJECT A SPATIAL TESTS FOR PREDICTING COMPREHENSIVE PERFORMANCE MEASURES

Henry H. Busciglio

Selection and Classification Technical Area
Manpower and Personnel Research Laboratory

U. S. Army Research Institute
for the Behavioral and Social Sciences

5001 Eisenhower Avenue
Alexandria, VA 22333-5600

February, 1990

The Usefulness of Project A Spatial Tests For Predicting Comprehensive Performance Measures

Introduction

The Army's Project A is a long-term, comprehensive effort to improve the selection and classification of enlisted personnel. One objective of this effort was to validate the Armed Services Vocational Aptitude Battery (ASVAB), the traditional instrument for assessing the general cognitive skills of potential soldiers. Previous analyses of Project A data (Campbell, 1988) demonstrated that the ASVAB is useful for predicting first tour performance, particularly such "can-do" criteria as General Soldiering Proficiency and Core (i.e., MOS-specific) Technical Proficiency. Therefore, the ASVAB serves as a baseline against which the marginal utility of other tests for selection and classification is judged.

Another objective of Project A was to develop and validate measures of abilities other than the general cognitive domain covered by ASVAB. For example, Project A staff members hypothesized that measures of spatial and perceptual-psychomotor abilities would account for criterion variance which was not predictable from ASVAB scores. In pursuit of this objective, a set of new Project A predictors was administered to 9,500 soldiers in 19 Military Occupational Specialties (MOS) in the 1985 Concurrent Validation phase.

In a previous analysis of the 1985 Concurrent Validity data, Busciglio (in preparation) found that Project A test scores substantially improved the prediction of many criteria, most notably three comprehensive measures of success: total score on written tests of school and job knowledge, General Soldiering Proficiency and Core (i.e., MOS-specific) Technical Proficiency. The statistical procedure used was a series of backward stepwise regressions in which the ASVAB subtests were entered into the equation in a block and nonsignificant subtests were removed, then the Project A tests were entered to determine the extent to which they were significant as incremental predictors.

The present analyses are similar except that spatial tests were entered first and then the ASVAB subtests. Two research topics were addressed:

1. How much criterion variance can the spatial tests alone account for, and how does this compare with the earlier results for ASVAB?

2. Which individual spatial tests will remain significant predictors of the criterion measures <u>after</u> the ASVAB subtests have been entered into the prediction equations?

Method

Subjects

Subjects were first-term enlisted personnel in the nine MOS for which hands-on criterion measures were collected as part of the 1985 Concurrent Validation phase of Project A. These MOS had been selected for comprehensive examination in Project A because they were judged to be representative of the entire population of entry-level Army MOS. The sample consists of individuals who had entered the Army between 1 July 1983 and 30 July 1984 and was drawn from thirteen posts in the continental United States as well as the U.S. Army in Europe (USAREUR). The number of subjects from each MOS, as well as the total sample size, is shown in Table 1.

Predictors

Predictors were the nine ASVAB subtests and the six Project A paper-and-pencil tests of spatial ability. Detailed information concerning the nature and development of the Project A predictors can be found in Peterson (1987). Table 2 presents a list of these predictors.

Criterion Measures

All criteria included in these analyses are comprehensive, "can-do" measures of proficiency in performing job duties, as described below.

Total Score on School and Job Knowledge Tests. School and job knowledge tests are written, multiple-choice measures of soldiers' technical knowledge pertinent to the various tasks performed in each MOS. Each school knowledge, or end-of-training test consists of 130-210 items, depending upon MOS (Davis, Davis, Joyner, & de Vera, 1987). The job knowledge tests are meant to measure knowledge of between 25 and 31 "critical tasks" and consist of 150-200 items, once again depending upon the particular MOS (Campbell, in preparation). Unlike items on the school knowledge tests, those on the job knowledge tests were selected during Project A to sample content broadly and to discriminate maximally among examinees. The total score is a unit-weighted composite of standard scores on the school and job knowledge tests.

General Soldiering Proficiency. General Soldiering and Core Technical Proficiency are the two "performance constructs" which the Project A staff created to account for the variance in "can-do" performance. General Soldiering Proficiency is a composite score on a variety of tasks common to many MOS (e.g., determining grid coordinates on maps, recognizing friendly vs. threat aircraft), as measured by written test items and hands-on tasks.

That is, the composite is an aggregate across measurement methods. (MOS 11B was not scored on this measure.)

Core (i.e., MOS-specific) Technical Proficiency. This is defined as soldiers' ability to perform the tasks that are at the "core" of each MOS (i.e., those that define the MOS), the composite score being an aggregate of scores on written test items and hands-on tasks.

Skill Qualification Test Score (SQT). These are paper-and-pencil tests of MOS-specific technical knowledge developed by the U.S. Army Training and Doctrine Command for periodic testing of soldiers in their MOS. This was the only criterion analyzed which was not developed in Project A. (MOS 91A was not scored on this measure.)

It should be stressed that the comprehensive measures above are not mutually exclusive. That is, written test scores were used in the computation of General Soldiering and Core Technical Proficiency, as well as the total score for written tests.

Procedure

Collection of Project A predictor and criterion data was part of the 1985 concurrent validation and occurred between 10 June and 13 November of that year. Scores on the ASVAB subtests and the Skill Qualification Test were obtained from archival data sources.

To answer the questions raised in the introduction, a series of backward stepwise multiple regression analyses were performed separately for each MOS. This procedure begins by entering all specified predictors into the equation as a block, and then removes nonsignificant predictors one by one, based upon their individual contributions to the overall R^2 with the criterion. This backward elimination procedure continues until all variables in the equation are individually significant at a certain probability level (the present analyses used .05).

The Army's selection and classification decisions are based on the ASVAB subtests. Because of this, restriction of range in the ASVAB scores (as well as any implicit range restriction in the measures which are correlated with them) would probably lead to underestimates of the actual criterion variance which would have been accounted for if ASVAB scores had not been used as a selection screen. To correct for this problem, matrices of estimated population (i.e., unrestricted) covariances among predictors and criteria were created, using the Lawley formula presented in Lord and Novick (1968; pp. 184-188), and used as input in these analyses. The R^2s reported below are in terms of this unrestricted population (the 1980 youth population, composed of individuals between the ages of 18 and 23) and have also been

adjusted for shrinkage, using the formula given by Wherry (1940).

Employing an SPSS Regression program with both the "Enter" and "Backward Stepwise" subcommands, the analyses proceeded in the following two stages:

1. SPATIAL ---> (SPATIAL)

The six spatial tests were entered as a block into each equation and the backward procedure determined which were individually significant.

2. (SPATIAL)+ASVAB ---> (SPATIAL+ASVAB)

Significant spatial tests were retained in a second stage of analysis, meaning that they were no longer subject to removal. The ASVAB subtests were entered as a block and retained only if they added individually significant variance to the R^2 obtained for the spatial tests alone.

The results of the present analysis were compared to those obtained by Busciglio (in preparation), which included the following stepwise procedures to assess the extent to which the spatial predictors could increment ASVAB:

1. ASVAB ---> (ASVAB)

The nine ASVAB subtests were entered as a block into each equation and the backward procedure determined which were individually significant.

2. (ASVAB)+SPATIAL ---> (ASVAB+SPATIAL)

Significant ASVAB subtests were retained in a second stage of analysis, meaning that they were no longer subject to removal. The spatial tests were entered as a block and retained only if they added individually significant variance to the R^2 obtained for ASVAB alone.

Results

First Research Topic

Table 3 shows R^2s and incremental R^2s for spatial and ASVAB tests as groups, comparing the results of the present analyses with those of the earlier investigation (Busciglio, in preparation). As can be seen, in many of the comparisons the spatial tests account for approximately as much criterion variance at Stage 1 of the analyses as do the ASVAB tests. The exceptions are shown in Table 4, which lists the criterion measures and MOS where the difference in R^2s for the two groups of tests equals or exceeds .05 (an arbitrary cutoff). As the

table shows, the ASVAB tests were superior to the spatial measures in all these instances, and were clearly better predictors of the two written criteria - Total Score on School and Job Knowledge Tests and the Skill Qualification Test. In contrast, the superiority of the ASVAB tests for predicting General Soldiering and Core Technical Proficiency, which are composites of written and hands-on measures, was limited to a smaller number of MOS.

There was some evidence that the superiority of ASVAB over spatial tests may be a function of MOS. As Table 4 indicates, the ASVAB subtests were superior predictors of most of the criterion measures in three MOS: 31C, 63B, and 91A.

Second Research Topic

Table 5 lists the individual spatial tests which attained significance in Stage 1 and notes with an asterisk those which remained significant after the ASVAB tests were entered (and nonsignificant ones removed) in Stage 2. The summary of these results below shows the number of times (out of a maximum of 34) that each spatial test was significant in Stage 1 and remained significant in Stage 2:

Spatial Test	Stage 1	Stage 2
Assembling Objects	23	21
Figural Reasoning	27	20
Map	34	18
Maze	6	0
Object Rotation	5	2
Orientation	16	1

As can be seen, Assembling Objects, Figural Reasoning, and Map were especially strong predictors of the criterion measures across MOS. Assembling Objects and Figural Reasoning remained significant in the vast majority of cases. The Map test, while the strongest of all predictors in Stage 1, remained significant a smaller percentage of the time. The other three spatial tests were much weaker overall and remained significant in only a very few cases.

Discussion

Findings pertinent to the first research topic indicated that the spatial tests measure abilities which are important across a wide variety of MOS and criteria, being comparable to ASVAB as predictors of the two composite criteria. Although the spatial tests were somewhat inferior as predictors of the two written performance measures, method variance might be responsible for these differences since the spatial tests, while in a written format, may not tap as much verbal or "scholastic"

ability as the ASVAB tests.

As reported above, there was some evidence to support the contention that the superiority of ASVAB over spatial tests for predicting comprehensive measures of performance may also be a function of MOS. The three enlisted jobs in the present sample where this seems to be the case - 31C, 63B, and 91A - are not easily distinguished from the others (see, for example, the titles of these jobs in Table 1). Thus, further research on the Project A predictors will be needed in order to determine what, if any, characteristics determine the extent to which job performance is differentially predictable from ASVAB vs. spatial measures.

Referring to the second research topic, three of the spatial tests - Assembling Objects, Figural Reasoning, and Map were very strong predictors across MOS and criteria and tended to remain significant after the ASVAB tests were entered and deleted in Stage 2. This supports the notion that these tests would lead to an overall improvement in the prediction of job performance if incorporated into the Army's selection composites. On the other hand, the usefulness of the Maze, Object Rotation, and Orientation tests seems to be much more limited to specific MOS and/or performance measure. Thus, their greatest utility would probably be in MOS-specific classification composites.

To interpret these results properly, it is important to note a number of methodological considerations. As stated earlier, individual written tests contributed variance to the total written score on school and job knowledge tests as well as to General Soldiering and Core Technical Proficiency. Also, possible differences in motivation due to differences in testing situations (i.e., ASVAB scores used for selection as opposed to Project A scores used "for research purposes only") may have impacted the results. That is, individuals may have responded more carefully, exerted more effort, etc., on the ASVAB subtests, thus making them more valid measures of abilities than the Project A tests. The final methodological concern has to do with the statistical analyses used. Stepwise regression procedures, while useful for empirically exploring alternative models, are especially susceptible to sampling error (cf. Cohen & Cohen, 1983). The samples used in the analyses were generally of sufficient size to make the degree of shrinkage in each <u>individual</u> equation relatively low, but the large <u>number</u> of equations computed here increases the probabilities that some ASVAB and Project A predictors are significant due to Type I errors. However, it should be noted that each spatial predictor can be significant a maximum of 34 times, across all criteria and MOS. At an alpha level of .05, a predictor may be expected to reach significance, by sampling error alone, between one and two times. As Table 5 clearly indicates, most of the Project A tests were significant far more often than this. Nevertheless, the

lack of opportunities at this point for cross-validation renders the results reported in this paper suggestive only.

REFERENCES

Busciglio, H.H. (in preparation). The incremental validity of spatial and perceptual-psychomotor tests in Project A's 1985 concurrent validation (ARI Technical Report in preparation). Alexandria, VA: U.S. Army Research Institute.

Campbell, C.H. (in preparation). Developing basic criterion scores for hands-on tests, job knowledge tests, and task rating scales (Draft of ARI Technical Report).

Campbell, J.P. (Ed.). (1988). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). Alexandria, VA: U.S. Army Research Institute.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/ correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, R.H., Davis, G.A., Joyner, J.N., & de Vera, M.V. (1987). Development and field test of job relevant knowledge tests for selected MOS (ARI Technical Report 757). Alexandria, VA: U.S. Army Research Institute.

Lord, P., & Novick, M. (1968). Statistical theory of mental test scores. Reading, MA: Addison-Wesley Publishing Co.

Peterson, N.G. (Ed.). (1987). Development and field test of the trial battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute.

Wherry, R.J. (1940). Appendix A in Stead, W.H., and Sharyle, C.P. Occupational counseling techniques.

Table 1

Subjects for the Present Analysis

MOS	Enlisted Job	N	SQT
11B	Infantry	491	444
13B	Cannon Crew	464	396
19E	Armor Crew	394	338
31C	Single Channel Radio Operator	289	248
63B	Light Wheel Vehicle Mechanic	478	409
64C (now 88M)	Motor Transport Operator	507	427
71L	Administrative Specialist	427	361
91A	Medical Specialist	392	0
95B	Military Police	597	545
TOTAL		4,039	3,168

Note. Sample sizes shown in SQT column were those available for the analysis of the Skill Qualification Test criteria, as described later in this paper.

Table 2

Predictors Used in the Analysis

ASVAB Subtests:	Spatial Ability Tests:
Arithmetic Reasoning	Assembling Objects
Auto/Shop Information	Figural Reasoning
Coding Speed	Map
Electronics Information	Maze
General Science	Object Rotation
Math Knowledge	Orientation
Mechanical Comprehension	
Number Operations	
Verbal (Paragraph Comprehension + Word Knowledge)	

Table 3

Comparison of R^2s for Spatial and ASVAB Tests

Analysis Stage Predictors	Present (1) (SPATIAL)	Present (2) (SPATIAL+ASVAB)	Previous (1) (ASVAB)	Previous (2) (ASVAB+SPATIAL)
Total Score on School and Job Knowledge Tests:				
11B	.59	.66	.59	.66
13B	.40	.45	.39	.45
19E	.57	.64	.58	.64
31C	.50	.61	.60	.62
63B	.50	.66	.63	.66
64C	.49	.59	.55	.59
71L	.55	.61	.54	.61
91A	.54	.70	.67	.69
95B	.58	.68	.62	.68
MEDIAN	.54	.64	.59	.64
General Soldiering Proficiency:				
11B	---	---	---	---
13B	.30	.33	.30	.33
19E	.47	.50	.44	.50
31C	.45	.51	.49	.52
63B	.31	.33	.28	.33
64C	.46	.53	.49	.53
71L	.44	.48	.41	.48
91A	.49	.56	.54	.56
95B	.56	.64	.59	.64
MEDIAN	.46	.51	.47	.51
Core Technical Proficiency:				
11B	.48	.53	.48	.53
13B	.16	.18	.15	.18
19E	.32	.39	.35	.38
31C	.38	.54	.54	.54
63B	.32	.50	.48	.50
64C	.28	.35	.32	.36
71L	.42	.49	.44	.49
91A	.45	.59	.58	.59
95B	.39	.45	.43	.46
MEDIAN	.38	.49	.44	.49

Table 3 (cont.)

Comparison of R^2s for Spatial and ASVAB Tests

Analysis Stage Predictors	Present (1) (SPATIAL)	Present (2) (SPATIAL+ASVAB)	Previous (1) (ASVAB)	Previous (2) (ASVAB+SPATIAL)
<u>Skill Qualification Test Score</u>:				
11B	.38	.47	.44	.46
13B	.19	.22	.19	.22
19E	.39	.44	.40	.44
31C	.44	.56	.55	.57
63B	.40	.56	.55	.56
64C	.38	.51	.50	.51
71L	.43	.59	.58	.60
91A	---	---	---	---
95B	.45	.60	.60	.61
MEDIAN	.40	.54	.53	.54

Table 4

Largest Differences in R^2s for Spatial and ASVAB Tests

Analysis Stage Predictors	Present (1) SPATIAL	Present (2) (SPATIAL)+ASVAB	Previous (1) ASVAB	Previous (2) (ASVAB)+SPATIAL
<u>Total Score on School and Job Knowledge Tests</u>:				
31C	.50	.61	.60	.62
63B	.50	.66	.63	.66
64C	.49	.59	.55	.59
91A	.54	.70	.67	.69
MEDIAN	.54	.64	.59	.64
<u>General Soldiering Proficiency</u>:				
91A	.49	.56	.54	.56
<u>Core Technical Proficiency</u>:				
31C	.38	.54	.54	.54
63B	.32	.50	.48	.50
91A	.45	.59	.58	.59
MEDIAN	.38	.49	.44	.49
<u>Skill Qualification Test Score</u>:				
11B	.38	.47	.44	.46
31C	.44	.56	.55	.57
63B	.40	.56	.55	.56
64C	.38	.51	.50	.51
71L	.43	.59	.58	.60
95B	.45	.60	.60	.61
MEDIAN	.40	.54	.53	.54

<u>Note</u>. Differences between R^2s for Spatial and ASVAB are greater than or equal to .05 (an arbitrary cutoff).

Table 5

Significant Spatial Predictors at Stage 1 of Analyses

<u>Total Score on School and Job Knowledge Tests</u>:						
11B		Fig Reas*	Map*			Orient
13B	Assem Objs*		Map*		Obj Rota	
19E	Assem Objs*	Fig Reas*	Map*			
31C		Fig Reas*	Map	Maze		
63B	Assem Objs*	Fig Reas*	Map*			Orient
64C	Assem Objs*	Fig Reas*	Map			
71L	Assem Objs*	Fig Reas*	Map*			
91A	Assem Objs*	Fig Reas*	Map			
95B	Assem Objs*	Fig Reas*	Map*			Orient
<u>General Soldiering Proficiency</u>:						
13B	Assem Objs*		Map*	Maze		
19E	Assem Objs*	Fig Reas*	Map*			
31C		Fig Reas*	Map	Maze		
63B	Assem Objs*		Map*			Orient
64C	Assem Objs*	Fig Reas*	Map		Obj Rota*	Orient*
71L	Assem Objs*	Fig Reas	Map		Obj Rota	Orient
91A	Assem Objs*	Fig Reas*	Map			Orient
95B	Assem Objs*	Fig Reas*	Map*			Orient
<u>Core Technical Proficiency</u>:						
11B		Fig Reas*	Map*			Orient
13B	Assem Objs*		Map*			
19E	Assem Objs*		Map*		Obj Rota*	
31C		Fig Reas	Map	Maze		
63B	Assem Objs*		Map			Orient
64C	Assem Objs*	Fig Reas*	Map			
71L	Assem Objs*	Fig Reas*	Map*		Obj Rota	
91A	Assem Objs	Fig Reas*	Map			
95B		Fig Reas*	Map*			Orient
<u>Skill Qualification Test Score</u>:						
11B		Fig Reas*	Map*			Orient
13B	Assem Objs*		Map*			
19E	Assem Objs*	Fig Reas	Map*			
31C		Fig Reas	Map	Maze		Orient
63B	Assem Objs	Fig Reas	Map			Orient
64C		Fig Reas	Map	Maze		Orient
71L		Fig Reas*	Map			
95B		Fig Reas	Map			Orient

* Remained significant after entry of ASVAB tests in Stage 2.
Assem Objs - Assembling Objects   Obj Rota - Object Rotation
Fig Reas - Figural Reasoning   Orient - Orientation

Working Paper

RS-WP-84-24

AN OUTLINE OF ISSUES PERTAINING TO EEO CONCERNS AND PROJECT A

John P. Campbell
HUMAN RESOURCES RESEARCH ORGANIZATION

October 1984

Reviewed by
Paul Rossmeissl
Selection and Classification
Technical Area

Paul Rossmeissl

Approved by
Newell K. Eaton
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

An Outline of Issues Pertaining to EEO Concerns and Project A

John P. Campbell

Since the passage of the 1964 Civil Rights Act and the beginning of litigation pertaining to discrimination in personnel selection, a great deal of effort has focused on carrying out research on the nature of employment discrimination, the explication of its meaning, and the development of potential remedies. There now is a very large literature on the topic, both conceptual and empirical, and virtually every professional doing business in personnel research or personnel management is familiar with it. The purpose of the memo is simply to remind everyone on Project A of the issues involved and to cite the relevant documentation and professional consensus, when it exists.

What follows is organized into three sections. First, some basic terms and their definitions are summarized. Second, the major questions and issues that comprise this topic are briefly outlined. For each issue relevant literature sources will be cited along with whatever professional consensus exists as to appropriate conclusions that can be drawn. The third section will attempt to summarize the relevance of these issues for Project A. The definitive treatment of these issues is to be found in the two volume report on "Ability Test: Uses, Consequences, and Controversies" published by the National Research Council and written by the Committee on Ability Testing, National Academy of Sciences (Wigdor & Garner, 1982).

SOME DEFINITIONS

The context for this discussion is that of personnel decision making; and more specifically a) selection and classification decisions for new employees and b) promotion decisions for incumbent employees. These two decisions are at the center of most of the literature on the topic. Each decision implies there is some available information (e.g., an interview or psychological test) upon which the decision is made and some means by which the accuracy or validity of the decision can be assessed.

Given this context, the major terms and concepts on which to be clear are the following.

VALIDITY

A professional consensus concerning the appropriate definition of validity is reflected in the professional literature and the policy statements of the American Psychological Association, the American Council on Measurement in Education, and the American Educational Research Association. In the context of personnel decision making the consensus is that a selection or promotion standard is valid if there is significant evidence to support the specific use for which it is intended. Significant evidence may be accumulated in several ways and the concepts of criterion related validation, content validation, and construct validation still constitute the most appropriate and professionally acceptable description of validation strategies. Consequently, while users are obligated to justify the use of particular decision making information with evidence, no one particular kind of evidence can be judged better than another on *a priori* grounds. The evidence that should be used depends upon the nature of the decision to be made, the context in which it must be made, the state of

previous research, and the feasibility of gathering various kinds of evidence. The definitive statement on these issues is the forthcoming "Joint Standards for the Development and Use of Psychological Tests" to be published by APA, ACME, and AERA in 1985.

DISCRIMINATION (IN SELECTION AND CLASSIFICATION)

Taken literally, the term simply refers to the act of making distinctions among job candidates. Consequently, discrimination could be carried out by choosing job candidates randomly, or on the basis of job qualifications, or on the basis of skin color, gender, hairlength, family background, etc.. There are hundreds of different ways that applicants could be hired and classified. Some of these methods are specifically forbidden by law. Others are deemed desirable or undesirable on other grounds. For example, hiring people by random selection from the available applicants is not forbidden by any current or previous statute. However, virtually everyone agrees that organizations should not have to use random selection if they don't want to. The notion that people should be selected (discriminated) on the basis of merit is still prevailing value judgment in U.S. society.

Much of the current consensus as to what kinds of discrimination are desirable and/or legal and which are not, has built up during the 20 years since the passage of the Civil Rights Act of 1964. Professional research, legislation, and litigation of individual cases during that period have contributed to this consensus. Developments during this 20 year period are thoroughly documented and described in other sources and will not be repeated here. The important points for our situation are that selection decisions (i.e., a discrimination between those that are hired and those

that are not hired) must be based on standards (i.e., predictors) that reflect legitimate job qualifications (i.e., merit). A job standard is legitimate if it is a valid predictor of job behavior, performance, or outcomes that represent a <u>business necessity</u>. This is another way of saying that selection measures (e.g., tests, interviews, job histories, educational histories, etc.) can be used to discriminate among applicants if they have significant validity for predicting a criterion that is relevant for the accomplishment of organizational goals. All legislation and all case law accept the principal of hiring on the basis of merit. Discrimination in hiring on the basis of race, religion, gender, ethnic background, or national origin is forbidden by Title VII. Discrimination on the basis of other variables unrelated to business necessity is also grounds for legal action on behalf of the individual who is denied employoment. Such actions can be taken under a variety of legal protections (e.g., State fair employment practice laws, Article XIV of the U.S. Constitution). The central issue is still whether or not the selection standard is a valid reflection of business necessity.

BIAS

The term bias is widely used with a multiplicity of meanings, and is difficult to define in any systematic way. Perhaps the best professional reference on the topic is the article by Cole (1981). What is presented here is simply an outline of her major distinctions. In general, the term refers to the fact that scores on a measure are influenced systematically by factors that are irrelevant for the stated purpose of measurement. The factors of specific interest for us are membership vs. non membership in one of the classes protected by Title VII. Within this context there are at least the following kinds of potential bias.

Selection Bias - Occurs when the selection decision itself systematically reflects class membership but when such class identification has nothing to do with job qualifications.

Test Bias - Refers to systematic group differences in total test scores that are irrelevant or unrelated to the purpose of measurement (e.g., group differences on a test of perceptual speed and accuracy that have nothing to do with the construct the test is intended to measure). Test bias could also occur even though group means were similar if the properties measured by the test were substantially different for the two groups.

Criterion Bias - Refers to systematic differences between groups that are unrelated to the components of job performance that the criterion is intended to measure.

DIFFERENTIAL PREDICTION

Differential prediction has a specific psychometric meaning and refers to a significant difference between the prediction functions for two groups when the task is to predict a measure of job performance from a selection measure. The function most often compared is the linear regression of the criterion on the predictor. Such functions can differ in terms of their slopes, intercepts, or standard errors of measurement. If the data are available, most professionals much refer to compare the complete regression functions rather than the zero order correlation coefficients. A significant difference between predictor/criterion correlations for the two groups is usually termed differential validity.

ADVERSE IMPACT

This is a legal term that refers to differences in selection ratios, or hiring rates, across groups when they are candidates for the same job. While the numerator in the two ratios is the number of people hired in each group, it is not always clear what the denominators should be. Are they a function of the actual applicant pool, a potential applicant pool, or the labor force itself? Typically, organizations have not been required by the Courts to meet unreasonable demands regarding their applicant pools.

FAIRNESS

The term fairness, or more specifically selection fairness, is also not a psychometric term but refers to whether or not the individual decisons made by the selection system are "fair" when held up against some set of community or societal standards. Previous attempts to give a psychometric definition to fairness were found wanting principally because the models which were proposed did not incorporate the necessary value judgments (Peterson & Novick, 1976).

QUESTIONS, ISSUES AND RESEARCH FINDINGS

Efforts to promote equal employment opportunity must deal systematically with a number of difficult questions and issues. For some of these questions and issues there is a reasonably large research literature upon which to base a course of action. For others there is not and much more work is needed. Listed below are the major questions and issues that are of current concern and a brief summary of the documentation that exists.

MEAN DIFFERENCES

As documented by the National Academy of Sciences panel on ability testing, the DoD's Survey of American Youth, and as well as many other sources (e.g., Jensen, 1980), there are significant mean difference on several key variables in the personnel decision making context. Probably the most consistent differences is the .75 - 1.0 standard deviation difference frequently observed between blacks and whites on standardized tests of verbal and arithmetic ability, with the white mean being higher than the black mean. Mean differences are frequently observed among other minority groups as well with the mean for the majority group usually being the highest. There are far fewer differences on ability tests between men and women; however, differences in favor of males are frequently found on measures of spatial relations (Sevy, 1983). Smaller differences are sometimes observed in favor of females on tests of verbal ability and in favor of males on arithmetic and mathematical ability.

The consistent mean differences that are found for majority-minority comparisons frequently do produce adverse impact for minority groups. The mean differences are not a product of unreliable measurement or flawed data collection techniques. However, the fact that they exist says nothing about their origin. The hundreds of studies that have been done do not permit any choice between an environment or heredity explanation. Another complication is that when the comparisons are made in a specific seclection situation the two groups are probably not sampled in the same way from their own reference population. For example, the individual expectations and perceptions that govern self selection may differ for the two groups and thus produce applicant pools that are different. Organizational

recruitment practices may also differ for the two groups as when the organization tries to implement affirmative action policies.

DIFFERENTIAL PREDICTION

There is now a large research literature related to differential prediction across racial and gender groups in both employment and educational settings. Again, the issue is whether there is a statistically significant difference in linear regression lines between groups. The research results, as documented in the National Academy of Sciences report, are consistent in showing that in both employment and educational settings there is seldom any significant differences in prediction functions between blacks and whites when an ability test is used to predict job performance. Further, the differences that do exist virtually always result in the job performance of blacks being overpredicted when a common regression line is used. This is because the regression line for blacks is almost always below that for whites. Consequently, it would virtually never be in the interests of the minority group to have a separate prediction equation.

How can there be both adverse impact for blacks vs. whites and no significant differential prediction? It is because there are also mean differences on the criterion and they are about what would be predicted from the mean differences on the predictor.

The above results could occur because both tests and criteria are baised or because both are correctly reflecting an average deficit in black performance. If in fact job performance criteria are generally unbiased, then the available research argues that selection tests are also unbiased. The relative mean differences are about what would be expected because of

regression effects. Since the research results comparing mean differences are reasonably consistent across different types of criteria, which at least appear to have varying susceptibility to bias (e.g., ratings vs. performance records), there is a stronger empirical argument for a lack of bias in performance criteria than there is to support its existence. After carefully reviewing all the evidence, the NAS panel concluded that well developed employment tests measuring general mental abilities are not biased against blacks.

These data do not, and should not, preclude the development of additional selection information that is also useful for predicting various aspects of job performance and which does not show adverse impact. For example, using tests of psychomotor ability to predict job proficiency or using interest/ biographical information to predict attrition were not addressed in the research examined by the NAS panel.

In contrast to the research on black/white differences, there is much less data regarding differential prediction across gender groups and the results are not so consistent. As yet, we have few consistent findings as to whether there are mean differences between males and females on predictors such as mathematical aptitude or spatial relations aptitude that are not reflected in relevant criterion differences.

A complicating factor in the interpretation of group differences in regression lines is that under certain conditions measurement error alone may produce such differences (Linn, 1984). For example, if the regression of the criterion true scores or the predictor true scores is the same for two groups, and the groups do indeed differ on the latent variable(s) that

produce the predictor/criterion correlation then measuring the predictor and criterion with less than perfectly reliable instruments will by itself produce a difference in intercepts. Consequently, it would be dangerous to interpret the over prediction of blacks as bias against whites.

TEST BIAS

The question of whether gender or minority group membership accounts for variance in total test scores that is independent of the characteristics that the test is intended to measure probably cannot be answered definitely. To do so would be analogous to saying that the construct validity of the test is completely determined, or that every possible alternative explanation has been determined. Measurement of individual abilities or behavior is not perfect, regardless of whether it is done by psychologists, economists, teachers, lawyers, blacks, or whites.

A more reasonable question is whether a test is biased in the way that it is used. We are concerned with the use of tests for selection and classification. Are tests biased for that purpose? If the comparison is between blacks and whites and the tests in question are well developed standardized tests of general cognitive abilities, the conclusion reached by NAS panel is that they are not. Their reasoning and the literature on which it is based is very carefully described in Volume I of their report. Since the question of whether there is test bias in selection is virtually the same as whether there is differential prediction of job performance criteria across groups the same evidence applies and the conclusions are the same.

ITEM BIAS

There has also been considerable research examining whether or not there is an interaction between item content or item type and group membership (e.g., Carroll and Horn, 1981; Jensen, 1980). While there are isolated examples of such interactions, there has been no general finding of item x group interactions (Carroll and Horn, 1981) within the broad array of items that are found on tests of general cognitive abilities. Consequently, it is not possible to attribute the source of the group differences to particular types of items. Another conclusion from these data is that the tests are apparently measuring the same construct within each group.

LINEAR VS. CURVILINEAR PREDICTION

An issue that is sometimes raised regarding the relationship of ability test scores to job performance is whether or not the regression of performance on ability is actually linear. An alternative hypothesis is that after a certain minimum level of ability is reached further increments in test scores do not mean further increments in job performance. Another way of stating it is to assert that beyond a certain point on the predictor continuum people can be selected at random and the average job performance for those selected would not be significantly lower than the performance of a group selected from the "top down" on the predictor. A related argument would be to assert that once an individual successfully completes a particular course of training or instruction, he or she is "qualified" for a particular job and no further distinctions need be made.

The evidence on this issue does not support the non-linear argument. For virtually all jobs the higher the scores on tests of general mental abilities, the higher the average job performance (Schmidt and Hunter,

1981). Also, while it is entirely possible that all the graduates of a particular course of instruction might reach an acceptable level of job performance, the average performance of the selected group (assuming a selection ratio of less than 1.0) will be higher if only the most able are selected.

Having said all this we must keep in mind that it is another question as to whether the gains in resultant performance from top down selection are worth the costs. For example, is the aggregate gain in performance worth the cost of testing? If adverse impact is designated as a cost then we can also ask whether the gains in average performance are worth the costs of adverse impact that are incurred.

CRITERION BIAS

Are there group differences on measures of job performance that are not accounted for by actual differences in job performance? If there is significant racial or gender bias in the criterion then empirical validity research will tend to emphasize predictors that are also biased. The principal conclusion that well developed tests of general mental abilities are not biased against blacks rests on the prior conclusion that the criterion measures used in personnel research incorporate no serious racial bias. The data on this question are reviewed by Arvey (1977) and Schmitt and Lappin (1980). While there is some data to suggest that raters tend to give higher performance ratings to members of their own race, there is also evidence suggesting the opposite (i.e., or reverse discrimination effect). However, the preponderance of the evidence suggests no significant difference.

RELEVANCE FOR PROJECT A

During the past 20 years a body of professional and public policy has been developed that governs the conduct of personnel decision making relative to equal employment opportunity. The relevant statute is the Civil Rights Act of 1964. The principal administrative policy guidelines are the Uniform Guidelines for Employment Selection. Professional policy is contained in the Joint Standards for the Use of Psychological Tests, published jointly by the American Psychological Association, the National Council of Measurement in Education, and the American Education Research Association, and the Principles for the Use of Tests in Personnel Selection (Rev. Ed.), published by the Division of Industrial and Organizational Psychology of the American Psychological Association.

Project A, as it was designed and as it is being carried out, will meet all the standards imposed by the federal government and by the policy documents of the relevant professions. In fact, it will go beyond these standards in an attempt to maximize both the effectiveness of the Army's personnel system and the employment opportunities of all individuals. Because of its comprehensive design, large sample sizes, and multiple methods of measurement, Project A is in a better position to deal with EEO issues than any previous selection/classification research study. It is incumbent on everyone to help the project realize its fullest potential in this regard. What follows is simply a list of the many different elements in Project A that have relevance for EEO issues.

A. Mean predictor and criterion differences for blacks/whites and males/females will be available for both new recruits and experienced job incumbents for several jobs and for a broad range of predictor

variables and criterion variables. One of the strengths of Project A in this regard is that an exhaustive search was made of all potential predictors and all potential criteria and a large sample of measures from each domain was selected without regard to the specific set of jobs to be studied. Consequently, we have an excellent opportunity to determine how the mean difference between groups varies across virtually the complete range of predictors and criteria that are available to personnel specialists. If there are types of variables for which adverse impact is significantly reduced we should be able to identify them. Maximizing this opportunity means that we should retain the widest possible variety of measures through as many data collections as possible.

B. In addition to looking at mean difference we can examine differential prediction across race and gender groups for a wide range of predictor/criterion combinations. If there are particular combinations which yield reasonable validity but low adverse impact we should be able to find them. Project A should go a long way toward a definitive answer regarding the extent to which it is potentially possible in employment settings to minimize adverse impact and maximize validity at the same time.

C. Within Project A the <u>form</u> of the prediction function can also be examined for the widest variety of predictor/criterion combinations ever studied in a single project. For example, we can ask whether there is in fact any particular class of criteria for which regression is nonlinear.

D. It seems generally agreed that the most powerful method for examining item bias across racial or gender groups is via item response theory techniques. For at least some MOS we should be able to use these techniques to calibrate items in both groups.

E. Project A is attempting to develop a utility metric by which the payoffs from classification under various conditions can be compared. If a reasonable metric can be developed then it will be possible to portray very systematically the tradeoffs between more or less adverse impact and the aggregate gain or loss in system utility. The comparative effects of random selection, random selection above a certain cutting score, etc., can also be simulated to portray the effects of a wide variety of value judgments and selection policy on the aggregate utility function.

The above elements will come together in essentially the following fashion. The first step is to develop criterion composites that maximize content and construct validity as comprehensive measures of enlisted job performance. That is, we want the best possible assessment of job performance without regard to how well each component might be predicted. At the same time, we must use all available evidence to insure that the criterion components do not incorporate racial or gender bias. The principal means for doing this are to show: 1) that any group differences that do exist are consistent across different methods and/or different groups of raters and 2) that a panel of experts does not judge the content of the criterion measures to be biased. The next step is to select predictor composites that maximize classification validity and minimize adverse impact. If adverse impact still exists but there is valid and

unbiased prediction of the performance criteria then a value judgment must be made as to the tradeoff between reducing adverse impact vs. reducing the aggregate performance of those selected and classified. If an appropriate performance utility metric can be developed it will greatly facilitate the meaningful description of the decision options.

A final warning is in order. Interpretations of group differences is fraught with psychometric pitfalls (cf Novick, 1984). Biased sampling from the reference population and regression effects due to unreliability can complicate the task of drawing conclusions from data. We must all proceed with caution.

SUMMARY

This memo is meant to be only a bare outline of EEO issues as they pertain to Project A. If you haven't done so already, every one should become familiar with the following basic sources:

- The 2-volume report of the National Academy of Sciences.
- The book on employment fairness by Richard Arvey.
- The October 1981 issue of the American Psychologist, which was a special issue devoted to testing and discrimination issues.

Selection and Classification Technical Area Working Paper RS-WP-88-5

SYNTHETIC VALIDATION PROJECT: PILOT TEST RESULTS

Wei Jing Chia
American Institutes for Research

Cyndi Owens-Kurtz & Norman Peterson
Personnel Decisions Research Institute

Philip L. Szenas
American Institutes for Research

March 1988

REVIEWED BY:

Jane M. Arabian
Research Psychologist

APPROVED BY:

Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Cleared by:

For Newell K.Eaton, Director

Manpower and Personnel Research Laboratory

**U. S. Army Research Institute
for the Behavioral and Social Sciences**
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Synthetic Validation Project: Pilot Test Results

Project Overview and Pilot Test Objectives

The two major objectives of the Army Synthetic Validation Project are to identify and evaluate procedures for:

- identifying an optimal composite of selection measures for any Army enlisted MOS and estimating the validity of this composite for predicting job performance

- setting a minimum qualifying score so as to assure a reasonable probability of successful job performance, as well as other appropriate cutting scores for other critical selection decisions (e.g., for selecting recruits with potential for outstanding performance).

Synthetic validation approaches typically begin with the identification of a set of job components that can be used to describe the population of jobs being studied. A prediction equation is derived for linking available selection tests to each component. Subject matter experts (SMEs) are asked to judge the importance of each component to overall job performance. Finally, the prediction equations for the various components are weighted according to the judgment weights and summed to obtain an equation for predicting overall performance for the job.

As a first step in developing procedures for obtaining overall performance prediction equations, we reviewed several alternative approaches to constructing a taxonomy of job components. Following the review, we proceeded with initial development of three job component models. The job descriptors used in the first model were task categories, those used in the second model were job activities or behaviors, and those used in the third model were individual attributes (e.g., abilities, traits, interests) required for successful job performance.

To date, we have completed a Pre-Test and a Pilot Test of the three models. Results of the Pre-Test were disseminated at the Scientific Advisory Committee meeting on November 20, 1987. Results of the Pilot Test are summarized in this report.

The goal of the Pilot Test was to assess the extent to which Army SMEs (NCOs and officers) can reliably identify the critical task categories, activities, and attribute requirements of Army jobs. Other issues included identifying possible subgroup differences in criticality ratings and identifying ways to fine-tune data collection procedures for each of the three models.

Method

Participants

Data were gathered from a total of 134 SMEs in three MOS -- Infantryman (11B), Vehicle/Generator Mechanic (63B), and Administrative Specialist (71L).

Two populations of judges participated in the job description workshops. The first was NCOs, officers, and civilians assigned to the Directorate of Training and Doctrine (DOTD) who help define doctrine and prepare training plans for each MOS. The second consisted of FORSCOM NCOs and officers who were supervising soldiers in these three MOS.

The DOTD data collection sites included Fort Benning (11B), Aberdeen Proving Ground (63B), and Fort Benjamin Harrison (71L). The workshops were conducted separately for NCOs and for officers.

Two FORSCOM installations were visited -- Fort Stewart and Fort Ord. These two installations were selected to ensure that our workshop participants included adequate samples of both mechanized infantrymen (Fort Stewart) and non-mechanized infantrymen (Fort Ord).

Tables 1a and 1b show the number of participants by rank and by command for each MOS. Additional demographic information can be found in Appendix A.

Procedure and Instruments

Each workshop began with the administrator giving an overview of project goals and a description of the specific objectives for the workshop. After the overview, participants completed judgment exercises that were organized into four sessions.

Session 1: Task Category and Job Activity Ratings and Task Matching. For each job component in the Task Category and Job Activity instruments (see Appendices B and C, respectively), participants provided importance, frequency, and difficulty ratings. They also matched each job component to up to 30 Project A job tasks. Following each matching exercise, participants completed a brief questionnaire which asked their opinions about the clarity and completeness of instructions and items of each component instrument. The order of presenting the two instruments was balanced across rank, command, site, and MOS.

Session 2: Attribute Importance and Validity Judgments and Attribute Ranking. Participants were presented with definitions of 31 attributes (cognitive abilities, perceptual-psychomotor abilities, temperaments, interests) and the five Project A performance factors (see Appendix D). They were asked to rate the importance or validity of each attribute for each performance factor. Participants also were asked to rank order the 31 attributes according to their importance or validity for overall

performance. After ranking the 31 attributes, participants completed a brief questionnaire which asked their opinions about the clarity and completeness of instructions and definitions as they were used in the attribute instrument.

Session 3: Attribute Importance Judgments for Job Components. From Session 1, the administrator selected up to 5 task categories and up to 5 job activities with the highest mean importance ratings. Participants were asked to rate the importance or validity of each of the 31 attributes for each of these 10 components.

Session 4: Group Consensus. The administrator also identified the five task categories and five job activities with the highest standard deviations. Workshop participants were asked to discuss their ratings of these job components. After this discussion, participants rerated the importance of these components.

Participants at Aberdeen Proving Ground completed sessions 1, 2, and 4, and participants at Ft. Benning completed sessions 1 and 2. Participants at others sites completed all four sessions, 1 and 2 in the morning and 3 and 4 in the afternoon.

Results: Task Category and Job Activity

Descriptive Statistics

One of the goals of the Pilot Testing was to assess whether judges can determine the relative importance, frequency, and difficulty of each job component for their MOS. As the three MOS have rather different job duties, it was critical that the judgments obtained for the task categories and job activities reflect the differences among the MOS.

For each task category and job activity, participants gave four judgments: (a) whether the component was part of the job, (b) importance, (c) frequency, and (d) difficulty. When a participant rated the component "not part of the job," a value of zero was entered for the participant's importance and frequency ratings.

Task Categories. Tables 2a-2c show, for each MOS, means and standard deviations for the importance, frequency, and difficulty ratings for the 97 task categories. Tables 3a-3c, show for each MOS, the task categories ordered by the mean importance ratings. The results in Tables 3a-3c show that tasks that are expected to be important for the MOS received the highest importance ratings. For 11B, items on combat, weapons, explosives, first aid, navigation, and communications received highest importance ratings. For 63B, items on using manuals, repairing, troubleshooting, vehicle operation, and general soldiering items were rated most important. For 71L, typing, document preparation and handling, and general soldiering items received the highest ratings. In all three MOS, a common set of general soldiering task categories such as navigate, give first aid, use maps, and fire individual weapons were among the twenty most important components.

Zero order correlations of the mean importance rating profiles between pairs of MOS were .66 for 11B and 63B, .48 for 11B and 71L, and .40 for 63B and 71L. These moderate correlations provide more support that our judges' ratings can be used to identify differences in critical tasks between the three MOS. The profile of importance weights was more similar between 11B and 63B than either MOS with 71L.

Although standard deviations for the mean importance ratings ranged from .13 to 2.24, the vast majority of them were less 2.0. Larger standard deviations were generally associated with components that received lower importance ratings.

Job Activities. Tables 4a-4c show, for each MOS, means and standard deviations for the importance, frequency, and difficulty ratings for the 53 job activities. Tables 5a-5c show, for each MOS, the activities ordered by the mean importance ratings. Again activities expected to be most important for each MOS were indeed rated most important. For 11B, items dealing with combat, weapons, physical activities, and following directions received the highest importance ratings. For 63B, items on operating hand tools and equipment, driving, troubleshooting, and following directions received the highest importance ratings. For 71L, items on operating keyboard, following or relaying directions, teamwork, and solving administrative problems received the highest importance ratings.

Pearson r's of the mean importance rating profiles between pairs of MOS were .61 for 11B and 63B, .35 for 11B and 71L, and .28 for 63B and 71L. These moderate correlations provide greater support that our judges' ratings can be used to identify differences in critical activities between the three MOS. The profile of importance weights was more similar between 11B and 63B than either MOS with 71L.

Although standard deviations for the mean importance ratings ranged from .43 to 2.20, most were less than 2.0. Again, larger standard deviations were generally associated with activities that received lower importance ratings.

Correlations among judgments. Table 6a shows the mean within-SME correlations between the importance, frequency, and difficulty ratings for the task categories that were rated "part of the job." In other words, these are correlations between non-missing ratings only. The correlations are reported separately for each MOS. The mean within-SME correlations between importance and frequency ratings were r=.48 for 11B, r=.57 for 63B, and r=.63 for 71L. The mean within-SME correlations between Difficulty and Importance or Frequency ratings ranged from r=-.09 to r=.25.

Table 6b shows the mean within-SME correlation between the importance, frequency, and difficulty ratings for the job activities that were rated "part of the job." The correlations are reported separately for each MOS. The mean within-SME correlations between importance and frequency ratings were r=.56 for 11B, r=.70 for 63B, and r=.70 for 71L. The mean within-SME

correlations between difficulty and importance or frequency ratings ranged from r=-.02 to r=.21.

These results suggest that, on both the task and activity instruments, there was some redundancy in importance and frequency information, but the difficulty ratings yielded rather different information. The difficulty ratings could serve as a tie-breaker for ordering components that are equal in importance or frequency.

Variance Component and Reliability Estimates

Variance component estimates provide a concise way of summarizing the differences among job descriptor means and describing the relative contribution of various conditions to variance in the ratings. Independent variables and the number of levels within each variable are as follows:

Condition	Number of Levels
Component	97 for Task Category; 53 for Job Activity
MOS	3 (11B, 63B, 71L)
Rank	3 (NCO, Officer, Civilian)
Command	2 (DOTD, FORSCOM)
Rater	134

As rater effects were nested within MOS, rank, and command, we ran two separate analysis of variance models. The first model included only rater and component. Estimates of the variance among all raters were obtained from this model. The second model included component, MOS, rank, and command, but not rater. Variance attributable to terms with MOS, rank, and/or command were subtracted from the overall rater variance estimates from the first model. The differences provided estimates of the variance due to rater nested within MOS, rank, and command.

Tables 7 to 12 provide variance component and reliability estimates for the task category and job activity importance ratings. Each table is based on a different variance component model:

(1) Component, MOS, Rank, Command, and Rater (Table 7)
(2) Component, MOS, Command, and Rater for each Rank (Tables 8a-8b)
(3) Component, MOS, Rank, and Rater for each Command (Tables 9a-9b)
(4) Component, Rank, Command, and Rater for each MOS (Tables 10a-10b)
(5) Component, Command, and Rater for each MOS and Rank (Tables 11a-11d)
(6) Component, Rank, and Rater for each MOS and Command (Tables 12a-12d)

From results in Tables 7 to 12, we underscore the following observations:

- the large variance component for both task category and job activity showed that, within MOS, judges reliably discriminated the important job components from the less important ones (see Tables 10a and 10b)

- the large variance component for the Component x MOS interaction suggested that weights were applied differentially for the job components across MOS (see Table 7)

Another important goal of the Pilot Testing was to assess whether SMEs can accurately and reliably apply the instruments in describing jobs.

Table 13a is a summary table of the average single-SME reliability estimates for both the task category and job activity instruments found in Tables 7 through 9. The table also shows the number of judges required to reach a target reliability of .90. The overall single-SME reliabilities of .58 for the task instrument and .45 for the activity instrument are respectable reliabilities.

Subgroup differences

Another issue in this research was the selection of SMEs for describing jobs. We were interested in determining whether judges of various ranks or from different assignments evaluated the jobs similarly. The research design included about equal numbers of NCOs and officers and about equal numbers of FORSCOM and DOTD judges.

The reliabilities in Table 13b and 13c show that officers were somewhat more reliable than NCOs for all three MOS and on both instruments. Overall reliabilities for DOTD and FORSCOM were generally comparable. DOTD judges were somewhat more reliable than FORSCOM judges for 11B and 63B. For 71L, FORSCOM judges were somewhat more reliable than DOTD judges.

Although there were small differences in reliabilities for different ranks, the rank ordering of task categories and activities was very consistent across ranks and commands. The correlation of mean task category importance rating profiles was .96 for 11B NCOs and officers, .95 for 63B NCOs and officers, and .95 for 71L NCOs and officers. Comparable correlations for the job activities were .92, .82, and .94 for 11B, 63B, and 71L, respectively. For the task categories, profile correlations for DOTD and FORSCOM were .95 for 11B, .91 for 63B, and .93 for 71L. Comparable correlations for the activity ratings were .95, .79, and .95, respectively.

In sum, reliabilities for both DOTD and FORSCOM were comparable, whereas officers were slightly more reliable than NCOs in their ratings. Also, NCOs and officers and DOTD and FORSCOM judges provided highly similar profiles.

Pilot Test v. Pre-Test Reliabilities

Corresponding Pre-Test reliabilities have also been included in the Table 13b and 13c. For 11B Pre-Test reliabilities were higher than Pilot Test reliabilities. For 71L, Pilot Test reliabilities were higher than Pre-Test values. For 63B, Pilot Test reliabilities were higher for the task instrument, but there were no reliability differences for the activity instrument.

Task Matching

The purpose of the Task Matching Exercise was twofold: (1) to provide an indication of the exhaustiveness of each job component model for describing jobs; and (2) to provide necessary linkages (via expert judgment) to enable the comparison between synthetic (inferred) validities and empirical validities obtained from Project A.

Results of the task matching exercise are reported in Tables 14a-14c. (Detailed results are reported in Appendix E.) The first column of Tables 14a-14c lists up to 30 Project A tasks for each MOS. For each task, SMEs identified up to three task categories and up to three activities that best matched it. The remaining columns show the job component number and the proportion who selected that component as one of their three task categories or three activities. Only job components selected by 50% or more of the SMEs are tabulated.

We found that: (1) SMEs could more readily match the job tasks to the task categories than to the job activities; (2) the SMEs used more task categories than job activities to cover the Project A tasks (e.g., 71Ls used 13 task categories and only 2 activities); (3) there were several tasks that were not covered by any of the task categories or activities; and (4) in several cases, content areas that were not covered by one model were covered by the other.

Tables 15a and 15b show the proportion of the important task categories and job activities, respectively, that were matched to the Project A tasks. It is clear that the number of important task categories that were matched to tasks were far greater than the number of important job activities that were matched to tasks. These results suggest that we may be unable to conduct comparisons between empirical validities and synthetic validities provided derived for activity-based job components.

The task matching exercise data suggest that:

- ambiguities still exist in several of the job component definitions (particulary in the use and meaning of the examples provided)
- the SMEs clearly were able to match task to task categories more easily with task categories than with job activities
- the definitions of some of the job activities may be too broad to link with specific tasks (e.g., the activity "Follow written directions" was mapped to both "Type a military letter" and "Put on field/pressure dressing" for 71L)
- neither model did a good job describing many of the 11B Project A combat tasks

Evaluation of Instruments

After the task matching exercise, a brief questionnaire was presented to judges for them to evaluate the task and the activity instruments. Copies of the evaluation questionnaires are included in Appendices B and C. Both ratings and open-ended comments were gathered from judges. One of the open-ended question asked judges to suggest components which should be deleted or added. Those who responded to the question had suggestions which were either too specific or too broad. There was also an insufficient number of comments to inform any decision. Other written-in general evaluations by judges were considered in revising the instruments. Means and distribution of judges' rating evaluation of the task instrument and the activity instrument are summarized in Tables 16 and 17, respectively. Here are some key observations:

- SMEs reported that the instructions on both instruments were clear

- although SMEs reported that the terms and definitions on both instruments were clear and easy to understand, SMEs found them somewhat clearer and easier to understand on the task instrument than on the activity instrument

- SMEs reported that the ratings were neither easy nor difficult to make

- SMEs were generally confident about the accuracy of their ratings

- except for 63B on the task instrument, both instruments provided, on the average, about 75% coverage of the jobs; for 63B, the task instrument provided 91% coverage

- subgroup comparisons between MOS, rank, and command did not reveal any large systematic differences

Conclusions: Task Category and Job Activity

The Pilot Test results confirmed that Army SMEs can reliably identify the critical task categories and activities of their jobs. Subgroup analyses between NCOs and officers and between DOTD and FORSCOM showed essentially no differences in subgroup ratings. There was no conclusive evidence that any one job component model is vastly superior to the others for describing jobs.

The results suggested ways the data collection could be improved. Based on the Pilot Test experience, we have made several changes for Phase I data collection procedures. These changes include:

- elimination of difficulty judgments for the task categories and activities

- separate importance ratings for Core Technical Proficiency, General Soldiering Proficiency, and overall job performance for each task category and activity

- elimination of the task matching exercise for task categories and activities instruments

- rewording task category and activity items with importance rating standard deviations of 2.0 or higher

- improved instructions, including better examples, for all three protocols.

- inclusion of a rating scale on the answer sheet

Results: Attributes

We turn now to a discussion of the results obtained from analyses of the Attribute ratings and rankings. We first discuss analyses of the responses the SMEs made on the evaluation questionnaire and then turn to the analyses of the ratings and rankings themselves.

Analyses of Evaluation Questionnaire

Description of Evaluation Questionnaire. The Attribute Model Evaluation Sheet had four sections: (1) performance definition, (2) attribute ranking, (3) attributes, (4) general evaluation (a complete copy is included in Appendix D). The performance definition and attribute ranking sections included ratings designed to assess the clarity of the performance definitions and the ranking task. The performance definition and attributes sections included open-ended questions designed to elicit comments concerning additions or deletions to the survey materials. The attributes section also contained an item asking participants to check any attributes that were difficult to judge or rank. In the general evaluation section, participants were asked to compare the judgment and ranking tasks on two dimensions: ease of use and amount of information provided. A third item in the general evaluation section asked participants which task they would choose to guide selection and placement of soldiers in MOS.

Evaluations were completed by 127 SMEs. The breakdown of participants by method, rater type, MOS, and command appears in Table 18.

Performance Definition and Ranking Task Ratings. SMEs indicated that the performance definition contained the right amount of detail and information (4.55 on a 7-point scale, with 4 representing "the right amount"). No significant differences were found when participants were subgrouped by method, rater type, MOS, or command. Table 19 presents the means and standard deviations by subgroup.

Overall, instructions for the ranking task were rated moderately clear (5.61 on a 7-point scale). Significant differences were found when participants were subgrouped by method and by rater type. SMEs who took

the Importance Task rated the ranking instructions as more clear than did participants who took the Validity Task ($F = 18.47$, $p < .001$). Civilians rated the ranking task instructions as less clear than did NCOs or Officers, and NCOs rated the instructions as less clear than did Officers ($F = 3.87$, $p = .023$). Means and standard deviations by subgroup are presented in Table 20.

The ranking task was rated neither easy nor difficult (4.27 on a 7-point scale). No significant differences were found for subgroups. Table 21 presents the means and standard deviations by subgroup.

Overall, participants were moderately confident of the accuracy of their rankings (4.92 on a 7-point scale). A significant difference in confidence ratings was found when participants were subgrouped by method. Those who took the Importance Task were more confident of their rankings than were those who took the Validity Task ($F = 5.75$, $p = .018$). Table 22 presents the means and standard deviations by subgroup.

No significant differences between MOS or command subgroups were found for the performance definition or ranking task ratings.

Correlations of the four evaluation ratings are presented in Table 23. Ratings of the amount of detail in the performance definition are unrelated to the three ranking task ratings. In the total sample, the three ranking task ratings are significantly correlated ($p < .001$). When the correlations are examined separately by method, the correlation between the ranking task clarity rating and the ranking task difficulty rating is not significant in the Importance Task subsample, but is significant ($p < .001$) in the Validity Task subsample.

Difficult Attributes. When asked to indicate at least two attributes that were difficult to judge or rank, 13 participants did not respond, 2 participants checked only one attribute, 51 participants checked two attributes, and 61 participants checked more than two attributes. The five attributes most frequently marked were Spatial Ability ($N = 35$), Closure ($N = 31$), Interest in Artistic Activities ($N = 25$), Interest in Protective Services ($N = 24$), and Interest in Science ($N = 23$). Frequencies for each of the 31 attributes are presented in Table 24.

Comments. Tables 25 through 28 present the performance definition and attribute comments. Most participants declined to respond when provided with opportunities to comment on the performance definition and the attributes. Several participants made comments pertaining to the attributes when asked about the performance definition, which indicates the instructions for the Evaluation Sheet should be clarified if it is to be used in the future. Six participants suggested the performance definition could be shortened. Twenty-one specific attributes were suggested for addition to the survey. Interest in Artistic Activities was most frequently nominated for deletion ($N = 16$).

General Evaluation. Table 29 contains frequencies and percentages for the general evaluation questions. A majority of participants indicated

that the judgment task was easier than the ranking task (50%) and provides more information (63%). When asked which task they would prefer to guide selection and placement of soldiers in MOS, 57% indicated a preference for the judgment task. Although "Both" and "Neither" were not response options, a few of the participants added them.

Summary of Evaluation Questionnaire Analyses. No alarming problems surfaced in these evaluations. This was the first time that we had attempted to use the ranking task, and were gratified that the raters found the ranking task doable and were moderately confident that their rankings were accurate.

Nearly the same attributes were identified as being difficult to rate in the pilot test as in the pre-test. These attribute definitions have received considerable attention, and it is our opinion that they can not be simplified further without losing their content or construct validity (that is, we could probably write simpler definitions, but they would no longer be defining the construct they are intended to define). These attributes (Spatial Ability, Closure, and several Interests) are probably just the most difficult subset to rate of this set of 31 attributes, and are not necessarily in need of further revision.

Perhaps the most interesting finding here is the preference expressed for using the validity or importance judgments (against the five performance dimensions) to "guide selection or placement of soldiers", rather than using the ranking of attributes (which appears to be less complex and takes less time for the rater) for that purpose.

For those interested, Appendix F contains a summary of responses to the four evaluation questions described in Tables 19-22.

Analyses of Attribute Ratings and Rankings of Importance and Validity

We turn now to the analyses of the judgments made about the Attributes by the SMEs. We first examine the impact of various factors such as MOS, Type of rater, and the Attributes themselves on the judgments made by the SMEs. We then look at the reliability of the judgments. Convergent and discriminant validity of the judgments are examined via multi-trait, multi-method logic. The relationships of the validity and importance ratings to the validity and importance rankings are also examined. Finally, we examine the mean values of the ratings and rankings in some detail.

Number of Available Raters. Tables 30 and 31 show the number of raters available by rating method (Importance or Validity), Rater Type (NCO, Officer, or Civilian), and MOS (11B, 63B, or 71L). Two things are noteworthy about these tables. First, the totals for the two tables sum to more than the total number of raters available (72 + 91 = 163 v.134). This occurs because some raters completed both validity and importance ratings. Second, the N's for civilians were extremely low for both methods, and the N for 63B was extremely low for the Importance method. These facts limited the analyses that could be completed, particularly the Analyses of Variance. We could complete the "full-blown" ANOVA for only the validity

ratings without civilians, if we wished to have a minimum cell size of 5, which we did.

Analyses of Variance of Validity Ratings and Rankings. Tables 32-37 show the results of ANOVAs of the validity ratings for each of the five performance areas and the validity rankings. There were two between-subjects factors (MOS and Type) and one within-subjects factor (Attribute).

Examination of these tables shows that there was a highly significant Attribute effect for all five performance areas and for the rankings. Thus, the attributes did receive different ratings and rankings when averaged across all three MOS and both raters. The Attribute x MOS effect was highly significant for the Core Technical performance area (Table 32) and for the ranking (for overall job performance, Table 37), and much less significant for the Physical Fitness/Military Bearing area (Table 36). It did not reach the .01 level of significance for the other three performance areas. This finding accords with our a priori expectations and with Project A empirical validity results, that is, there are validity differences for attributes with respect to the Core Technical part of soldier's jobs, but not with respect to the "common" parts of a soldier's job--which are represented by the other four performance areas.

There are small but statistically significant Attribute x Type effects for all five performance areas, but not for ranking. This finding indicates that there may be sufficient rater type differences (NCO v. Officer only, in these analyses) to warrant including both types of raters in operational ratings. Or we may wish to study the exact nature of the differences in the ratings given by NCOs v. Officers and choose to use only one type in the operational setting.

Calculation of Inter-Rater Agreement Reliabilities. We computed inter-rater agreement reliability coefficients for validity ratings and ranking (Table 38) and for importance ratings and ranking (Table 39). The coefficients were computed using the output from a one-way repeated measures ANOVA to estimate mean squares for attributes and for error; rater main effect was thus included as part of the error term. Note that we calculated a reliability for each MOS x Performance area x Rater type cell. With one exception (63B NCOs and Officers: Importance, in Table 39) we only calculated reliabilities where the cell N was at least 10 in order to achieve minimum stability for mean square estimates.

As Tables 38 and 39 show, the ratings and rankings are reasonably reliable for both importance and validity. The smallest reliability in either table is .73 (for k, or number of raters, = 15), and almost half (39 of 84) are .90 or greater. There does not appear to be a practically significant difference in reliability between the two methods, but if there is a slight edge, it probably favors the importance method. However, there were generally fewer data points available for the importance method and this slight edge might disappear if more data were available.

There is not enough information in these tables to compare Officer and NCO ratings for all MOS and both methods, but what information is there indicates that Officers generally agree better than do NCO's.

Multi-Trait, Multi-Method Analyses. If one thinks of MOS as traits and the Importance and Validity ratings as two different methods, then these data can be analyzed via the logic of the multi-trait, multi-method approach. Tables 40-46 show the results of such analyses for the five performance areas separately, the rankings, and all five areas at once (that is, using the full 155 item profile that is made up of all Attribute (31) x performance area (5) ratings). These striking results are summarized below:

Summary of Multi-trait, Multi-method Analyses

Means of the Correlations displayed in Convergent Validity (Bet. Method, Within MOS) and Discriminant Validity (Bet. Method, Between MOS and Within Method, Bet. MOS) Portions of Tables 40-46: Core Technical, General Soldiering, Effort/Leadership, Personal Discipline, and Physical Fitness/Military Bearing Performance Areas, Overall Ranking, and All Performance Areas.

	CORE	GEN	EFF	DIS	FIT	RNK	ALL
Bet. Method, Within MOS	.89	.84	.87	.89	.94	.92	.91
Bet. Method, Between MOS	.42	.83	.84	.85	.92	.46	.79
Within Method, Bet. MOS	.42	.83	.84	.86	.92	.45	.79
* Within Importance	.38	.80	.78	.81	.93	.36	.81
* Within Validity	.45	.84	.90	.90	.92	.54	.75

Note that the ratings for the Core Technical Performance Area and the Overall Rankings show very high mean coefficients for convergent validity, as do all the other areas. Note, further, that the mean coefficients for discriminant validity are forty points lower for these two ratings than for the other four performance areas (.42 and .46 versus .83-.92 for the between method, between MOS mean coefficients and .42 and .45 v. .83-.92 for the within method, between MOS mean coefficients), indicating that different profiles of attributes are obtained for the three MOS for Core Technical and for Overall Ranking, but not for the other four performance areas--that is, discriminant validity exists for these two but not for the other four. As we stated earlier, these are precisely the results we would expect based on prior research conducted by Project A regarding the validity relationships between predictors and these five performance areas.

The last two rows of the summary show that there is slightly better discriminant validity for Importance ratings than for Validity ratings (.36 v. .45 for Core Technical and .36 v. 54 for Rankings). Again, we urge

caution in overinterpreting these relatively small differences, given the size of the present samples.

Relationship of Mean Importance and Mean Validity Ratings to Rankings. Tables 47-49 show the correlations of the mean importance and mean validity rating profiles (that is, the mean rating given to each of the thirty-one attributes) with the mean importance and mean validity ranking profiles (that is, the mean ranking assigned to each of the thirty-one attributes) for the three MOS. A separate correlation has been computed for each performance area. These tables represent a sort of rough "policy-capturing" of the rankings of the attributes for overall performance in the MOS. That is, if the mean rating profile for Core Technical correlates very highly with the mean ranking profile and the mean rating profile for Fitness/Bearing correlates very lowly with the mean ranking profile, then we might conclude that the Core Technical performance area is "more important" for overall performance than is Fitness/Bearing--"more important" in the sense that the attributes rated as highly and lowly valid for Core Technical are also ranked similarly highly and lowly valid for Overall Performance.

Examination of these three tables shows that:

- For 11B, Core Technical and General Soldiering are nearly equal and higher than the other three--but the other three are still relatively high.

- For 63B, Core Technical is much greater than any of the others, General Soldiering is moderately high, and the others are very low.

- For 71L, Core Technical is extremely high, General Soldiering, Effort/Leadership, and Personal Discipline are about 30 points less, but still fairly high, and Fitness/Bearing is zero.

- In all three tables, the validity and importance ratings show the same pattern of relationship with the rankings, but the validity ratings consistently correlate lower with the rankings than do the importance ratings. This could indicate that validity ratings are less susceptible to an overall "halo" effect than are the importance ratings.

Mean Importance and Mean Validity Rating and Ranking Profiles. Mean importance and validity ratings for the 31 attributes were computed by MOS for each of the five job areas (Tables 50-54 and 56-60). We then compared these mean rating profiles across MOS, but within job performance area. Attributes for which the maximum difference in means between any pair of MOS was equal to or greater than 2.5 were noted. (Although this size of mean difference was somewhat arbitrarily chosen, the standard error of difference for almost all mean comparisons for these data is less than 1.2. Therefore the value of 2.5 will almost always represent a difference greater than two standard errors of difference.) It was expected that the mean ratings would differ by MOS for the Core Technical Proficiency job area but not for other job areas. This reasoning was based on the notion,

previously mentioned, that General Soldiering, Effort and Leadership, Personal Discipline, and Physical Fitness/Military Bearing are common across MOS and should consequently be related in the same way to the attributes for all MOS, while Core Technical Proficiency is specific to the MOS and, within each MOS, would be related uniquely to the 31 attributes.

The mean importance ratings of the 31 attributes did not correspond closely to the expected pattern. Core Technical Proficiency did have 12 attributes showing a maximum difference at or above the 2.5 cut-off: spatial ability, closure, mechanical comprehension, precision, movement judgment, physical strength, physical endurance, balance and flexibility, involvement in athletics, interest in using tools and machines, interest in rugged activities, and interest in protective services. However, the other four areas each had at least two attributes (rather than none) that also met this criteria. For General Soldiering, 2 attributes met the maximum difference cut: spatial ability and involvement in athletics. For Effort and Leadership, 8 attributes met the maximum difference cut: number ability, eye-limb coordination, precision, movement judgment, physical strength, balance and flexibility, interest in rugged activities, and interest in protective services. For Personal Discipline, 8 attributes met the maximum difference cut: number ability, precision, movement judgment, physical strength, physical endurance, balance and flexibility, involvement in athletics, and interest in rugged activities. For Physical Fitness/Military Bearing, 3 attributes met the maximum difference cut: dominance/confidence, interest in protective services, and interest in leadership.

Validity ratings of the 31 attributes for the five job areas closely followed the expected pattern. Nine attributes met the maximum difference cut for Core Technical Proficiency: mechanical comprehension, movement judgment, physical strength, physical endurance, balance and flexibility, involvement in athletics, interest in using tools and machines, interest in rugged activities, and interest in protective services. No attributes met the maximum difference cut for General Soldiering, Effort and Leadership, or Personal Discipline. Two attributes met the maximum difference cut for Physical Fitness/Military Bearing: work orientation, and interest in efficiency and organization.

SMEs also rank ordered the 31 attributes in terms of importance or validity for overall job performance. The mean ranks are shown in Tables 55 and 61, respectively. The five attributes given the highest mean importance rankings for 11B were physical endurance, physical strength, verbal ability, reasoning, and mental information processing (in order). For 63B they were mechanical comprehension, interest in using tools and machines, interest in technical activities, hand and finger dexterity, and work orientation. For 71L the five highest were verbal ability, reasoning, mental information processing, work orientation, and memory.

The five attributes given the highest mean validity rankings for 11B were physical endurance, physical strength, reasoning, memory and mental information processing. For 63B, the top five mean validity rankings went to reasoning, mechanical comprehension, mental information processing,

interest in using tools and machines, and memory. The five attributes given the highest mean validity rankings for 71L were verbal ability, hand and finger dexterity, reasoning, mental information processing, and memory.

For 11B and 71L, the top five validity rankings closely parallel the top five rankings on importance. The 63B validity and importance rankings are dissimilar, however. No ready explanation for the 63B differences presents itself.

Conclusions: Attributes

We were pleased by the results of the pilot test. On the whole, we think the results show that Army personnel understand the instructions and other materials used in the three rating tasks (importance ratings of attributes for the five job performance areas, validity ratings of attributes for the five job performance areas, and rankings of validity or importance of attributes for overall job performance). They can use these materials to make the necessary judgments and they agree on those judgments to a practically useful degree. The pattern of the judged relationships between the attributes and the five job performance areas, as revealed in the various analyses detailed above, corresponds strikingly to the pattern that we expected to find and to the pattern found in Project A research.

We had also hoped that the pilot test might point clearly to either the validity or the importance methods as the preferred method. Alas, we did not obtain such clear-cut results. Both methods produced reliable results and the methods converged on similar results. We did not achieve sufficient N's to make some of the comparisons between methods that were planned, but we think that the results we do have strongly indicate that either method will produce meaningful, reliable results. Therefore, we have opted to pursue the use of the validity method in Phase I, primarily because we also will use the validity method when we collect judgments from psychologists (linking the attributes to task categories and job activities). It seems best to use the same method for both sets of judges so that when it comes time to compare the two sets of judgments we will have removed at least one source of ambiguity in the comparison.

The results show that Army judges can make reliable judgments that link the 31 attributes to the five job performance areas. We do not have judgment data showing how well they can link the 31 attributes to more specific entities--such as the task categories or job activities. This is the judgment that we will ask the psychologists to make. Therefore, we think it would be prudent to ask Army judges to link the attributes to a subset of these more specific job descriptors during Phase I.

For Phase I data collection then, we will eliminate the collection of importance ratings and rankings, and use only the validity method. Improved instructions, including a standard script, will be prepared. The validity ratings will be collected for two of the five job performance areas, Core Technical and General Soldiering, but not for the other three areas. Instead, five "common" job task category items (rated most

important across the three MOS) and five "specific" job task category items (rated most important within the MOS, but not "common") will be used.

General Conclusions

The results for all three job component models show that Army personnel understood the procedures and materials for judging the importance and/or validity of individual component for job performance. The results also show that there was good agreement among the judges on those judgments. However, there was no conclusive evidence for favoring one job component model over the others.

Specific changes to each job component model data collection protocol have been discussed. We will also be making several changes in all three instruments primarily to standardize the way terms are used across all three exercises. We will:

- provide a consistent description of the target job; judges will be instructed to focus on soldiers with 18 months on-the-job experience beyond Basic Training and Advanced Individual Training and told to consider the range of possible job duties within the MOS when making their judgments

- improve instructions, including better examples, for all three protocols.

Table 1a Number of Participants by Rank and MOS

RANK FREQUENCY PERCENT ROW PCT COL PCT	MOS 11B	63B	71L	TOTAL
NCO	36 26.87 46.75 60.00	21 15.67 27.27 60.00	20 14.93 25.97 51.28	77 57.46
OFFICER	17 12.69 36.96 28.33	11 8.21 23.91 31.43	18 13.43 39.13 46.15	46 34.33
CIVILIAN	7 5.22 63.64 11.67	3 2.24 27.27 8.57	1 0.75 9.09 2.56	11 8.21
TOTAL	60 44.78	35 26.12	39 29.10	134 100.00

Table 1b Number of Participants by Command and MOS

COMMAND FREQUENCY PERCENT ROW PCT COL PCT	MOS 11B	63B	71L	TOTAL
FORSCOM	34 25.37 51.52 56.67	14 10.45 21.21 40.00	18 13.43 27.27 46.15	66 49.25
DOTD	26 19.40 38.24 43.33	21 15.67 30.88 60.00	21 15.67 30.88 53.85	68 50.75
TOTAL	60 44.78	35 26.12	39 29.10	134 100.00

Table 2a
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 11B: Infantryman

	Task Category	Importance N	Importance MEAN	Importance S.D.	Frequency N	Frequency MEAN	Frequency S.D.	Difficulty N	Difficulty MEAN	Difficulty S.D.
1	Perform operator maintenance checks an	60	3.82	1.62	60	2.22	1.01	52	1.52	0.58
2	Perform operator checks and services o	60	4.63	0.69	60	2.77	0.50	59	1.39	0.53
3	Inspect mechanical systems	60	1.73	2.06	60	0.93	1.16	26	1.85	0.67
4	Repair weapons	60	1.47	2.00	60	0.58	0.85	22	1.86	0.77
5	Repair mechanical systems	60	1.63	2.06	60	0.77	1.01	24	1.96	0.55
6	Troubleshoot mechanical systems	60	1.10	1.87	60	0.53	0.93	17	2.06	0.56
7	Troubleshoot weapons	60	2.43	2.24	60	1.27	1.33	34	1.82	0.76
8	Install electronic components	60	3.10	1.86	60	1.63	1.06	46	1.52	0.55
9	Inspect electrical systems	60	1.07	1.78	60	0.53	0.95	17	2.00	0.61
10	Inspect electronic systems	60	0.95	1.68	60	0.47	0.87	16	2.00	0.63
11	Repair electrical systems	60	0.82	1.58	60	0.38	0.80	14	1.93	0.73
12	Repair electronic components	60	0.47	1.24	60	0.22	0.61	8	2.25	0.71
13	Troubleshoot electrical systems	60	0.80	1.53	60	0.43	0.89	14	2.00	0.78
14	Troubleshoot electronic components	60	0.52	1.31	60	0.28	0.74	9	2.33	0.50
15	Pack and load materials	60	2.38	1.80	60	1.15	0.94	43	1.49	0.63
16	Prepare parachutes	60	0.53	1.42	60	0.25	0.70	8	2.38	0.52
17	Prepare equipment and supplies for ai	60	0.93	1.64	60	0.40	0.72	17	2.12	0.78
18	Operate power excavating equipment	60	0.30	1.15	60	0.17	0.67	4	1.75	0.50
19	Operate wheeled vehicles	60	3.42	1.54	60	1.83	0.91	55	1.35	0.52
20	Operate track vehicles	60	3.78	1.45	60	2.07	0.92	56	1.52	0.54
21	Operate boats	60	1.03	1.72	60	0.43	0.74	19	1.47	0.61
22	Operate lifting, loading, and grading	60	0.12	0.58	60	0.07	0.31	3	1.33	0.58
23	Paint	60	1.25	1.34	60	0.98	0.95	40	1.07	0.27
24	Install wire and cables	60	1.92	1.80	60	0.98	0.95	37	1.43	0.60
25	Repair plastic and fiberglass	60	0.13	0.70	60	0.12	0.56	3	2.00	0.00
26	Repair metal	60	0.20	0.94	60	0.12	0.56	3	2.00	1.00
27	Assemble steel structures	60	0.20	0.86	60	0.12	0.49	4	2.00	0.82
28	Install pipe assemblies	60	0.02	0.13	60	0.02	0.13	1	1.00	.
29	Construct wooden buildings and other	60	0.25	0.84	60	0.20	0.63	7	1.29	0.49
30	Construct masonry buildings and struc	60	0.17	0.78	60	0.08	0.42	3	1.33	0.58
31	Operate gas and electric powered equi	60	0.97	1.40	60	0.55	0.81	24	1.58	0.58
32	Select, layout and clean medical or d	60	0.00	0.00	60	0.00	0.00	0	.	.
33	Use audiovisual equipment	60	0.67	1.36	60	0.42	0.79	17	1.59	0.62
34	Reproduce printed material	60	0.52	1.02	60	0.43	0.85	16	1.13	0.34
35	Operate electronic equipment	60	2.27	1.94	60	1.23	1.11	38	1.79	0.78
36	Operate radar	60	0.17	0.91	60	0.08	0.46	2	3.00	0.00
37	Operate computer hardware	60	0.47	1.21	60	0.27	0.69	10	2.60	0.52
38	Cook	60	0.32	1.05	60	0.18	0.60	6	1.33	0.52
39	Perform medical laboratory procedures	60	0.00	0.00	60	0.00	0.00	0	.	.
40	Conduct land surveys	60	1.53	2.13	60	0.83	1.20	22	2.05	0.72
41	Provide medical or dental treatment	60	0.42	1.39	60	0.18	0.65	5	1.40	0.55
42	Sketch maps, overlays, or range cards	60	3.97	1.44	60	2.08	0.91	55	1.80	0.65
43	Produce technical drawings	60	0.22	0.90	60	0.12	0.49	4	1.75	0.96
44	Draw maps and overlays	60	0.80	1.69	60	0.42	0.93	12	1.92	0.79
45	Draw illustrations	60	1.63	1.85	59	0.80	1.00	29	1.76	0.79
46	Type	60	0.53	1.17	60	0.33	0.71	13	1.85	0.69
47	Prepare technical forms and documents	60	1.07	1.59	60	0.63	0.99	22	1.86	0.83
48	Record, file, and dispatch informatio	60	0.90	1.56	60	0.48	0.83	18	1.89	0.68
49	Receive, store, and issue supplies, e	60	0.82	1.61	60	0.48	0.97	14	1.93	0.83
50	Use hand and arm signals	59	3.95	1.28	59	2.29	0.89	56	1.23	0.47
51	Read technical manuals, field manuals	60	3.72	1.37	60	2.12	0.94	56	1.50	0.63
52	Use maps	60	4.33	1.10	60	2.27	0.86	58	2.02	0.66

Table 2a continued
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 11B: Infantryman

	Task Category	Importance N	Importance MEAN	Importance S.D.	Frequency N	Frequency MEAN	Frequency S.D.	Difficulty N	Difficulty MEAN	Difficulty S.D.
53	Send and receive radio messages	60	4.28	0.87	60	2.18	0.70	60	1.75	0.57
54	Give oral reports	60	4.32	0.98	60	2.35	0.71	59	1.56	0.65
55	Receive clients, patients, guests	60	0.17	0.91	60	0.08	0.46	2	1.50	0.71
56	Give directions and instructions	60	3.07	1.74	60	1.70	1.05	49	1.51	0.62
57	Write documents and correspondence	60	0.58	1.31	60	0.30	0.72	11	2.45	0.82
58	Write and deliver presentations	60	0.28	1.03	60	0.13	0.47	5	2.40	0.89
59	Interview	60	0.65	1.45	60	0.27	0.63	11	1.64	0.50
60	Provide counseling and other interper	60	1.17	1.93	60	0.60	1.06	17	2.18	0.81
61	Decode data	60	3.30	1.83	60	1.40	0.94	48	2.10	0.81
62	Analyze electronic signals	60	0.43	1.25	60	0.17	0.49	8	2.25	0.71
63	Analyze weather conditions	60	0.65	1.41	60	0.33	0.73	13	1.69	0.75
64	Order equipment and supplies	60	1.02	1.70	60	0.50	0.81	19	1.68	0.75
65	Estimate time and cost of maintenance	60	0.20	0.94	60	0.13	0.60	3	2.00	1.00
66	Plan placement or use of tactical equ	60	1.72	2.19	60	0.80	1.13	24	2.13	0.80
67	Translate foreign languages	60	0.28	1.09	60	0.10	0.44	4	3.00	0.00
68	Analyze intelligence data	60	0.57	1.51	60	0.20	0.58	8	2.50	0.76
69	Control money	60	0.33	1.11	60	0.20	0.68	6	1.83	0.98
70	Determine firing data for indirect fi	60	0.90	1.71	60	0.42	0.85	14	2.07	0.73
71	Compute statistics or other mathemati	60	0.27	0.95	60	0.15	0.61	5	2.00	1.00
72	Provide programming and data processi	60	0.02	0.13	60	0.02	0.13	1	1.00	.
73	Control air traffic	60	0.27	1.04	60	0.07	0.25	4	2.75	0.50
74	Use hand grenades	60	4.37	1.07	60	1.80	0.86	58	1.28	0.49
75	Protect against NBC hazards	60	4.65	0.61	60	2.23	0.72	60	1.70	0.70
76	Handle demolitions or mines	60	3.83	1.65	60	1.52	0.85	53	2.02	0.72
77	Engage in hand-to-hand combat	60	4.05	1.06	60	1.58	0.79	59	1.61	0.70
78	Fire individual weapons	60	4.87	0.43	60	2.48	0.72	60	1.58	0.67
79	Control individuals and crowds	60	3.05	1.82	60	1.27	0.88	49	1.53	0.68
80	Customs and laws of war	60	3.77	1.41	60	1.58	0.87	57	1.54	0.63
81	Navigate	60	4.37	1.22	60	2.08	0.87	57	2.07	0.73
82	Survive in the field	60	4.73	0.52	60	2.52	0.68	60	1.57	0.70
83	Load and unload field artillery or ta	60	0.53	1.50	60	0.28	0.83	7	1.57	0.79
84	Fire heavy direct fire weapons (e.g.,	60	1.87	2.17	60	0.87	1.10	27	2.00	0.68
85	Prepare heavy weapons for tactical us	60	0.75	1.64	60	0.38	0.85	11	2.09	0.83
86	Place and camouflage tactical equipme	60	3.82	1.58	60	1.83	0.99	54	1.69	0.72
87	Fire indirect fire weapons (e.g., fie	60	0.62	1.45	60	0.32	0.77	10	2.30	0.67
88	Give first aid	60	4.52	0.68	60	2.02	0.72	60	1.68	0.62
89	Detect and identify targets	60	4.30	1.08	60	2.12	0.88	58	1.69	0.65
90	Plan, organize, monitor	60	1.38	1.91	60	0.63	0.92	23	1.91	0.73
91	Clarify roles, provide feedback	60	1.40	1.95	60	0.80	1.22	22	1.95	0.72
92	Provide information	60	2.52	2.12	60	1.30	1.20	38	1.71	0.65
93	Recognize, reward	60	1.03	1.71	60	0.58	1.00	18	1.94	0.80
94	Support	60	1.60	2.00	60	0.87	1.16	26	1.85	0.73
95	Train, develop	60	1.25	1.96	60	0.65	1.05	19	2.16	0.76
96	Discipline, punish	60	0.97	1.67	60	0.50	0.91	16	2.00	0.63
97	Act as model	60	2.92	2.12	60	1.78	1.33	42	2.02	0.81
98	Conduct tactical operations	60	1.22	1.98	60	0.55	0.91	18	2.50	0.62

Table 2b
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 63B: Vehicle/Generator Mechanic

	Task Category	Importance			Frequency			Difficulty		
		N	MEAN	S.D.	N	MEAN	S.D.	N	MEAN	S.D.
1	Perform operator maintenance checks an	35	4.69	0.58	35	2.63	0.60	35	1.54	0.66
2	Perform operator checks and services o	35	3.63	1.57	35	1.63	0.91	33	1.45	0.56
3	Inspect mechanical systems	35	4.46	1.22	35	2.43	0.85	33	1.88	0.70
4	Repair weapons	35	0.40	1.19	35	0.14	0.43	4	2.00	0.82
5	Repair mechanical systems	35	4.46	1.22	35	2.57	0.81	33	1.76	0.66
6	Troubleshoot mechanical systems	35	4.43	1.27	34	2.29	0.87	32	2.13	0.66
7	Troubleshoot weapons	35	0.63	1.33	35	0.29	0.57	8	1.88	0.83
8	Install electronic components	35	1.03	1.65	35	0.46	0.78	11	1.91	0.54
9	Inspect electrical systems	35	4.14	1.24	35	2.17	0.82	33	2.33	0.65
10	Inspect electronic systems	35	0.69	1.55	35	0.37	0.84	6	2.50	0.55
11	Repair electrical systems	35	4.17	0.95	35	2.11	0.68	35	1.97	0.66
12	Repair electronic components	35	1.83	1.99	35	1.00	1.11	17	2.06	0.75
13	Troubleshoot electrical systems	35	4.09	1.29	35	2.17	0.86	33	2.24	0.71
14	Troubleshoot electronic components	35	1.37	2.03	35	0.77	1.19	12	2.17	0.72
15	Pack and load materials	35	2.06	1.63	35	1.11	0.93	26	1.46	0.58
16	Prepare parachutes	35	0.00	0.00	35	0.00	0.00	0	.	.
17	Prepare equipment and supplies for ai	35	0.14	0.49	35	0.09	0.28	3	2.33	0.58
18	Operate power excavating equipment	35	0.09	0.51	35	0.06	0.34	1	1.00	.
19	Operate wheeled vehicles	35	4.31	0.87	35	2.51	0.74	35	1.37	0.60
20	Operate track vehicles	35	1.77	1.96	35	1.11	1.25	18	1.56	0.51
21	Operate boats	35	0.00	0.00	35	0.00	0.00	0	.	.
22	Operate lifting, loading, and grading	35	1.31	1.81	35	0.86	1.19	15	1.67	0.49
23	Paint	35	2.09	1.56	35	1.34	0.87	30	1.10	0.40
24	Install wire and cables	35	0.71	1.15	35	0.46	0.70	12	1.25	0.62
25	Repair plastic and fiberglass	35	0.20	0.47	35	0.23	0.60	6	2.33	0.82
26	Repair metal	35	1.51	1.54	35	0.80	0.87	20	1.90	0.85
27	Assemble steel structures	35	0.00	0.00	35	0.00	0.00	0	.	.
28	Install pipe assemblies	35	0.11	0.53	35	0.09	0.37	2	1.50	0.71
29	Construct wooden buildings and other	35	0.06	0.24	35	0.11	0.53	2	1.00	0.00
30	Construct masonry buildings and struc	35	0.00	0.00	35	0.00	0.00	0	.	.
31	Operate gas and electric powered equi	35	2.60	1.68	35	1.34	0.97	28	1.57	0.63
32	Select, layout and clean medical or d	35	0.00	0.00	35	0.00	0.00	0	.	.
33	Use audiovisual equipment	35	0.26	0.85	35	0.14	0.43	4	1.00	0.00
34	Reproduce printed material	35	0.11	0.32	35	0.11	0.32	4	1.00	0.00
35	Operate electronic equipment	34	0.65	1.10	34	0.29	0.46	10	2.10	0.57
36	Operate radar	35	0.00	0.00	35	0.00	0.00	0	.	.
37	Operate computer hardware	35	0.40	0.98	35	0.26	0.66	6	1.83	0.75
38	Cook	35	0.00	0.00	35	0.00	0.00	0	.	.
39	Perform medical laboratory procedures	35	0.00	0.00	35	0.00	0.00	0	.	.
40	Conduct land surveys	35	0.51	1.27	35	0.26	0.56	7	2.00	0.82
41	Provide medical or dental treatment	35	0.14	0.85	35	0.03	0.17	1	2.00	.
42	Sketch maps, overlays, or range cards	35	1.34	1.70	35	0.57	0.70	16	1.75	0.68
43	Produce technical drawings	35	0.00	0.00	35	0.00	0.00	0	.	.
44	Draw maps and overlays	35	0.03	0.17	35	0.03	0.17	1	1.00	.
45	Draw illustrations	35	0.26	0.61	35	0.17	0.38	6	1.50	0.55
46	Type	35	0.37	0.88	35	0.20	0.47	6	1.50	0.55
47	Prepare technical forms and documents	35	2.14	2.05	35	1.37	1.31	20	1.75	0.64
48	Record, file, and dispatch informatio	35	1.54	1.80	35	0.83	1.04	17	1.65	0.70
49	Receive, store, and issue supplies, e	35	1.06	1.63	35	0.60	0.95	13	1.46	0.52
50	Use hand and arm signals	35	2.83	1.72	35	1.57	1.04	30	1.30	0.60
51	Read technical manuals, field manuals	35	4.80	0.53	35	2.69	0.58	35	1.83	0.66
52	Use maps	35	3.14	1.44	35	1.31	0.83	32	1.81	0.74

Table 2b continued
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 63B: Vehicle/Generator Mechanic

	Task Category	Importance			Frequency			Difficulty		
		N	MEAN	S.D.	N	MEAN	S.D.	N	MEAN	S.D.
53	Send and receive radio messages	35	2.11	1.81	35	1.06	0.97	24	1.92	0.72
54	Give oral reports	35	2.57	1.63	35	1.17	0.79	30	1.53	0.73
55	Receive clients, patients, guests	35	0.03	0.17	35	0.03	0.17	1	3.00	.
56	Give directions and instructions	35	2.91	1.38	35	1.63	0.84	32	1.28	0.52
57	Write documents and correspondence	35	0.09	0.51	35	0.06	0.34	1	2.00	.
58	Write and deliver presentations	35	0.00	0.00	35	0.00	0.00	0	.	.
59	Interview	35	0.51	1.34	35	0.31	0.83	5	1.40	0.55
60	Provide counseling and other interper	35	0.09	0.51	35	0.06	0.34	1	2.00	.
61	Decode data	35	0.54	1.34	35	0.26	0.66	6	2.33	0.82
62	Analyze electronic signals	35	0.00	0.00	35	0.00	0.00	0	.	.
63	Analyze weather conditions	35	0.09	0.37	35	0.06	0.24	2	1.00	0.00
64	Order equipment and supplies	35	1.91	2.02	35	0.97	1.10	18	1.67	0.59
65	Estimate time and cost of maintenance	35	0.51	1.22	35	0.26	0.61	7	1.57	0.79
66	Plan placement or use of tactical equ	35	0.91	1.65	35	0.40	0.74	10	1.80	0.79
67	Translate foreign languages	35	0.00	0.00	35	0.00	0.00	0	.	.
68	Analyze intelligence data	35	0.00	0.00	35	0.00	0.00	0	.	.
69	Control money	35	0.00	0.00	35	0.00	0.00	0	.	.
70	Determine firing data for indirect fi	35	0.14	0.55	35	0.09	0.28	3	1.33	0.58
71	Compute statistics or other mathemati	35	0.17	0.62	35	0.11	0.40	3	2.00	1.00
72	Provide programming and data processi	35	0.06	0.34	35	0.03	0.17	1	1.00	.
73	Control air traffic	35	0.00	0.00	35	0.00	0.00	0	.	.
74	Use hand grenades	35	2.74	1.67	35	1.17	0.79	31	1.29	0.53
75	Protect against NBC hazards	35	4.09	1.20	35	1.69	0.87	35	1.69	0.72
76	Handle demolitions or mines	35	1.20	1.69	35	0.49	0.66	14	1.57	0.51
77	Engage in hand-to-hand combat	35	2.40	1.79	35	1.06	0.91	26	1.58	0.64
78	Fire individual weapons	35	3.97	1.10	35	1.60	0.69	35	1.46	0.61
79	Control individuals and crowds	35	1.37	1.61	35	0.77	0.91	19	1.53	0.61
80	Customs and laws of war	35	2.86	1.57	35	1.17	0.79	31	1.55	0.68
81	Navigate	35	2.94	1.59	35	1.26	0.89	30	1.77	0.82
82	Survive in the field	35	3.69	1.59	35	1.51	0.92	32	1.50	0.67
83	Load and unload field artillery or ta	35	0.00	0.00	35	0.00	0.00	0	.	.
84	Fire heavy direct fire weapons (e.g.,	35	0.14	0.85	35	0.03	0.17	1	2.00	.
85	Prepare heavy weapons for tactical us	35	0.00	0.00	35	0.00	0.00	0	.	.
86	Place and camouflage tactical equipme	35	2.80	1.83	35	1.31	0.99	27	1.41	0.64
87	Fire indirect fire weapons (e.g., fie	35	0.00	0.00	35	0.00	0.00	0	.	.
88	Give first aid	35	3.80	1.28	35	1.54	0.92	34	1.68	0.68
89	Detect and identify targets	35	2.54	1.72	35	1.09	0.85	27	1.70	0.78
90	Plan, organize, monitor	35	1.14	1.75	35	0.51	0.82	12	2.08	0.79
91	Clarify roles, provide feedback	35	0.77	1.50	35	0.37	0.77	8	2.25	0.89
92	Provide information	35	1.94	1.89	35	0.97	1.01	20	1.55	0.69
93	Recognize, reward	35	0.71	1.32	35	0.34	0.68	9	1.67	0.87
94	Support	35	1.17	1.67	35	0.51	0.78	13	1.85	0.80
95	Train, develop	35	0.94	1.61	35	0.46	0.82	10	2.00	0.82
96	Discipline, punish	35	0.60	1.29	35	0.29	0.67	7	2.14	0.90
97	Act as model	35	3.40	1.59	35	2.20	1.08	31	1.71	0.74
98	Conduct tactical operations	35	0.43	1.14	35	0.23	0.65	5	2.20	0.84

Table 2c
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 71L: Administrative Specialist

	Task Category	Importance			Frequency			Difficulty		
		N	MEAN	S.D.	N	MEAN	S.D.	N	MEAN	S.D.
1	Perform operator maintenance checks an	39	1.87	1.59	39	0.77	0.63	26	1.50	0.58
2	Perform operator checks and services o	39	2.77	1.77	39	1.15	0.78	32	1.56	0.56
3	Inspect mechanical systems	39	0.33	0.96	39	0.15	0.43	5	1.20	0.45
4	Repair weapons	39	0.21	0.73	39	0.10	0.38	3	2.00	1.00
5	Repair mechanical systems	39	0.08	0.48	39	0.03	0.16	1	1.00	.
6	Troubleshoot mechanical systems	39	0.18	0.79	39	0.08	0.35	2	2.00	0.00
7	Troubleshoot weapons	39	0.79	1.56	39	0.28	0.56	9	1.89	0.33
8	Install electronic components	39	1.54	1.73	39	0.72	0.83	20	1.85	0.49
9	Inspect electrical systems	39	0.44	1.12	39	0.21	0.52	6	1.67	0.52
10	Inspect electronic systems	39	0.18	0.79	39	0.08	0.35	2	2.00	0.00
11	Repair electrical systems	39	0.13	0.80	39	0.05	0.32	1	2.00	.
12	Repair electronic components	39	0.15	0.81	39	0.08	0.35	2	2.50	0.71
13	Troubleshoot electrical systems	39	0.03	0.16	39	0.03	0.16	1	1.00	.
14	Troubleshoot electronic components	39	0.00	0.00	39	0.00	0.00	0	.	.
15	Pack and load materials	39	0.77	1.13	39	0.46	0.64	15	1.40	0.51
16	Prepare parachutes	39	0.13	0.80	39	0.08	0.48	1	2.00	.
17	Prepare equipment and supplies for ai	39	0.10	0.64	39	0.05	0.32	1	1.00	.
18	Operate power excavating equipment	39	0.10	0.64	39	0.05	0.32	1	1.00	.
19	Operate wheeled vehicles	39	2.23	1.35	39	1.18	0.72	33	1.24	0.44
20	Operate track vehicles	39	0.15	0.71	39	0.10	0.50	2	1.50	0.71
21	Operate boats	39	0.00	0.00	39	0.00	0.00	0	.	.
22	Operate lifting, loading, and grading	39	0.13	0.66	39	0.10	0.50	2	2.00	0.00
23	Paint	39	0.51	0.60	39	0.46	0.51	18	1.00	0.00
24	Install wire and cables	39	0.56	1.14	39	0.28	0.56	9	1.33	0.50
25	Repair plastic and fiberglass	39	0.00	0.00	39	0.00	0.00	0	.	.
26	Repair metal	39	0.03	0.16	39	0.03	0.16	1	1.00	.
27	Assemble steel structures	39	0.00	0.00	39	0.00	0.00	0	.	.
28	Install pipe assemblies	39	0.00	0.00	39	0.00	0.00	0	.	.
29	Construct wooden buildings and other	39	0.00	0.00	39	0.00	0.00	0	.	.
30	Construct masonry buildings and struc	39	0.00	0.00	39	0.00	0.00	0	.	.
31	Operate gas and electric powered equi	39	1.13	1.36	39	0.54	0.55	20	1.80	0.83
32	Select, layout and clean medical or d	39	0.13	0.80	39	0.08	0.48	1	3.00	.
33	Use audiovisual equipment	39	0.62	1.21	39	0.36	0.67	10	1.60	0.97
34	Reproduce printed material	39	3.00	1.34	39	2.10	1.02	36	1.47	0.61
35	Operate electronic equipment	39	2.13	1.75	39	1.23	1.09	25	2.04	0.54
36	Operate radar	39	0.13	0.80	39	0.08	0.48	1	3.00	.
37	Operate computer hardware	39	3.74	1.04	39	2.28	0.79	38	2.18	0.65
38	Cook	39	0.15	0.81	39	0.10	0.50	2	1.50	0.71
39	Perform medical laboratory procedures	39	0.03	0.16	39	0.03	0.16	1	1.00	.
40	Conduct land surveys	39	0.67	1.24	39	0.33	0.58	11	2.09	0.70
41	Provide medical or dental treatment	39	0.13	0.80	39	0.05	0.32	1	2.00	.
42	Sketch maps, overlays, or range cards	39	1.08	1.42	39	0.51	0.64	17	1.82	0.73
43	Produce technical drawings	39	0.13	0.52	39	0.10	0.38	3	2.00	1.00
44	Draw maps and overlays	39	0.28	0.94	39	0.15	0.49	4	2.00	0.82
45	Draw illustrations	39	0.64	1.09	39	0.41	0.68	13	2.00	0.71
46	Type	39	4.82	0.45	39	2.95	0.22	39	1.85	0.63
47	Prepare technical forms and documents	39	4.51	0.64	39	2.74	0.44	39	1.82	0.64
48	Record, file, and dispatch informatio	39	4.51	0.68	39	2.79	0.41	39	1.74	0.64
49	Receive, store, and issue supplies, e	39	3.36	1.50	39	1.92	0.98	35	1.54	0.56
50	Use hand and arm signals	39	1.08	1.33	39	0.62	0.71	20	1.30	0.47
51	Read technical manuals, field manuals	39	3.77	1.44	39	2.23	0.78	38	1.84	0.72
52	Use maps	39	2.31	1.54	39	1.08	0.77	32	1.84	0.72

Table 2c continued
Task Category Importance, Frequency, and Difficulty Rating Descriptive Statistics
For 71L: Administrative Specialist

	Task Category	Importance N	Importance MEAN	Importance S.D.	Frequency N	Frequency MEAN	Frequency S.D.	Difficulty N	Difficulty MEAN	Difficulty S.D.
53	Send and receive radio messages	39	2.00	1.62	39	0.90	0.75	28	1.71	0.60
54	Give oral reports	39	2.56	1.45	39	1.23	0.74	34	1.47	0.51
55	Receive clients, patients, guests	39	2.28	1.78	39	1.54	1.17	28	1.39	0.63
56	Give directions and instructions	39	3.28	1.21	39	2.03	0.84	38	1.58	0.60
57	Write documents and correspondence	39	3.21	1.63	39	1.87	0.95	34	2.03	0.63
58	Write and deliver presentations	39	0.90	1.29	39	0.44	0.68	14	2.14	0.77
59	Interview	39	1.38	1.68	39	0.72	0.94	17	1.65	0.70
60	Provide counseling and other interper	39	0.74	1.35	39	0.44	0.79	11	1.55	0.69
61	Decode data	39	0.74	1.37	39	0.33	0.62	11	2.18	0.60
62	Analyze electronic signals	39	0.03	0.16	39	0.03	0.16	1	2.00	.
63	Analyze weather conditions	39	0.03	0.16	39	0.03	0.16	1	2.00	.
64	Order equipment and supplies	39	2.15	1.53	39	1.23	0.93	28	1.32	0.48
65	Estimate time and cost of maintenance	39	0.13	0.57	39	0.05	0.22	2	3.00	0.00
66	Plan placement or use of tactical equ	39	0.46	1.00	39	0.26	0.50	9	1.56	0.53
67	Translate foreign languages	39	0.15	0.67	39	0.10	0.38	3	2.33	1.15
68	Analyze intelligence data	39	0.21	0.80	39	0.08	0.27	3	1.67	0.58
69	Control money	39	0.82	1.65	39	0.41	0.85	9	1.56	0.73
70	Determine firing data for indirect fi	39	0.36	1.09	39	0.13	0.34	5	2.00	0.71
71	Compute statistics or other mathemati	39	0.79	1.26	39	0.46	0.72	13	2.15	0.69
72	Provide programming and data processi	39	0.87	1.49	39	0.56	0.94	12	2.17	0.83
73	Control air traffic	39	0.00	0.00	39	0.00	0.00	0	.	.
74	Use hand grenades	39	2.03	1.63	39	0.87	0.66	29	1.38	0.49
75	Protect against NBC hazards	39	3.13	1.69	39	1.08	0.62	35	1.74	0.66
76	Handle demolitions or mines	39	1.13	1.70	39	0.44	0.64	15	2.00	0.76
77	Engage in hand-to-hand combat	39	1.72	1.57	39	0.74	0.64	26	1.54	0.65
78	Fire individual weapons	39	3.15	1.73	39	1.10	0.72	34	1.79	0.59
79	Control individuals and crowds	39	0.90	1.23	39	0.44	0.55	16	1.50	0.63
80	Customs and laws of war	38	2.16	1.53	38	0.87	0.53	30	1.37	0.56
81	Navigate	39	2.21	1.64	39	0.87	0.61	30	2.03	0.76
82	Survive in the field	39	3.03	1.72	39	1.10	0.64	35	1.74	0.70
83	Load and unload field artillery or ta	39	0.15	0.81	39	0.10	0.50	2	2.00	0.00
84	Fire heavy direct fire weapons (e.g.,	39	0.13	0.66	39	0.10	0.50	2	2.00	0.00
85	Prepare heavy weapons for tactical us	39	0.10	0.64	39	0.05	0.32	1	2.00	.
86	Place and camouflage tactical equipme	39	1.33	1.66	39	0.56	0.68	18	1.61	0.61
87	Fire indirect fire weapons (e.g., fie	39	0.21	0.80	39	0.08	0.27	3	2.33	0.58
88	Give first aid	39	3.18	1.52	39	1.10	0.45	37	1.68	0.58
89	Detect and identify targets	39	1.95	1.56	39	0.85	0.59	29	1.90	0.77
90	Plan, organize, monitor	39	1.62	1.80	39	0.92	1.04	21	1.62	0.74
91	Clarify roles, provide feedback	38	1.13	1.63	38	0.74	1.08	15	1.93	0.80
92	Provide information	38	1.87	1.71	39	1.15	1.14	24	1.46	0.66
93	Recognize, reward	39	1.15	1.57	39	0.64	0.90	16	1.50	0.82
94	Support	39	1.72	1.82	39	0.92	1.01	21	1.62	0.74
95	Train, develop	39	1.49	1.78	39	0.77	0.99	19	1.84	0.76
96	Discipline, punish	39	1.03	1.55	39	0.51	0.79	14	1.93	0.83
97	Act as model	39	2.92	1.87	39	1.87	1.26	30	1.87	0.82
98	Conduct tactical operations	39	0.46	1.21	39	0.26	0.72	6	1.50	0.84

Table 3a
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 11B: Infantryman

No.	Task Category	N	MEAN	S.D.
78	Fire individual weapons	60	4.87	0.43
82	Survive in the field	60	4.73	0.52
75	Protect against NBC hazards	60	4.65	0.61
2	Perform operator checks and services	60	4.63	0.69
88	Give first aid	60	4.52	0.68
74	Use hand grenades	60	4.37	1.07
81	Navigate	60	4.37	1.22
52	Use maps	60	4.33	1.10
54	Give oral reports	60	4.32	0.98
89	Detect and identify targets	60	4.30	1.08
53	Send and receive radio messages	60	4.28	0.87
77	Engage in hand-to-hand combat	60	4.05	1.06
42	Sketch maps, overlays, or range cards	60	3.97	1.44
50	Use hand and arm signals	59	3.95	1.28
76	Handle demolitions or mines	60	3.83	1.65
1	Perform operator maintenance checks an	60	3.82	1.62
86	Place and camouflage tactical equipme	60	3.82	1.58
20	Operate track vehicles	60	3.78	1.45
80	Customs and laws of war	60	3.77	1.41
51	Read technical manuals, field manuals	60	3.72	1.37
19	Operate wheeled vehicles	60	3.42	1.54
61	Decode data	60	3.30	1.83
8	Install electronic components	60	3.10	1.86
56	Give directions and instructions	60	3.07	1.74
79	Control individuals and crowds	60	3.05	1.82
97	Act as model	60	2.92	2.12
92	Provide information	60	2.52	2.12
7	Troubleshoot weapons	60	2.43	2.24
15	Pack and load materials	60	2.38	1.80
35	Operate electronic equipment	60	2.27	1.94
24	Install wire and cables	60	1.92	1.80
84	Fire heavy direct fire weapons (e.g.,	60	1.87	2.17
3	Inspect mechanical systems	60	1.73	2.06
66	Plan placement or use of tactical equ	60	1.72	2.19
5	Repair mechanical systems	60	1.63	2.06
45	Draw illustrations	60	1.63	1.85
94	Support	60	1.60	2.00
40	Conduct land surveys	60	1.53	2.13
4	Repair weapons	60	1.47	2.00
91	Clarify roles, provide feedback	60	1.40	1.95
90	Plan, organize, monitor	60	1.38	1.91
23	Paint	60	1.25	1.34
95	Train, develop	60	1.25	1.96
98	Conduct tactical operations	60	1.22	1.98
60	Provide counseling and other interper	60	1.17	1.93
6	Troubleshoot mechanical systems	60	1.10	1.87
9	Inspect electrical systems	60	1.07	1.78
47	Prepare technical forms and documents	60	1.07	1.59
21	Operate boats	60	1.03	1.72
93	Recognize, reward	60	1.03	1.71
64	Order equipment and supplies	60	1.02	1.70
31	Operate gas and electric powered equi	60	0.97	1.40

Table 3a continued
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 11B: Infantryman

No.	Task Category	N	MEAN	S.D.
96	Discipline, punish	60	0.97	1.67
10	Inspect electronic systems	60	0.95	1.68
17	Prepare equipment and supplies for ai	60	0.93	1.64
48	Record, file, and dispatch informatio	60	0.90	1.56
70	Determine firing data for indirect fi	60	0.90	1.71
11	Repair electrical systems	60	0.82	1.58
49	Receive, store, and issue supplies, e	60	0.82	1.61
13	Troubleshoot electrical systems	60	0.80	1.53
44	Draw maps and overlays	60	0.80	1.69
85	Prepare heavy weapons for tactical us	60	0.75	1.64
33	Use audiovisual equipment	60	0.67	1.36
59	Interview	60	0.65	1.45
63	Analyze weather conditions	60	0.65	1.41
87	Fire indirect fire weapons (e.g., fie	60	0.62	1.45
57	Write documents and correspondence	60	0.58	1.31
68	Analyze intelligence data	60	0.57	1.51
16	Prepare parachutes	60	0.53	1.42
46	Type	60	0.53	1.17
83	Load and unload field artillery or ta	60	0.53	1.50
14	Troubleshoot electronic components	60	0.52	1.31
34	Reproduce printed material	60	0.52	1.02
12	Repair electronic components	60	0.47	1.24
37	Operate computer hardware	60	0.47	1.21
62	Analyze electronic signals	60	0.43	1.25
41	Provide medical or dental treatment	60	0.42	1.39
69	Control money	60	0.33	1.11
38	Cook	60	0.32	1.05
18	Operate power excavating equipment	60	0.30	1.15
58	Write and deliver presentations	60	0.28	1.03
67	Translate foreign languages	60	0.28	1.09
71	Compute statistics or other mathemati	60	0.27	0.95
73	Control air traffic	60	0.27	1.04
29	Construct wooden buildings and other	60	0.25	0.84
43	Produce technical drawings	60	0.22	0.90
26	Repair metal	60	0.20	0.94
27	Assemble steel structures	60	0.20	0.86
65	Estimate time and cost of maintenance	60	0.20	0.94
30	Construct masonry buildings and struc	60	0.17	0.78
36	Operate radar	60	0.17	0.91
55	Receive clients, patients, guests	60	0.17	0.91
25	Repair plastic and fiberglass	60	0.13	0.70
22	Operate lifting, loading, and grading	60	0.12	0.58
28	Install pipe assemblies	60	0.02	0.13
72	Provide programming and data processi	60	0.02	0.13
32	Select, layout and clean medical or d	60	0.00	0.00
39	Perform medical laboratory procedures	60	0.00	0.00

Table 3b
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 63B: Vehicle/Generator Mechanic

No.	Task Category	N	MEAN	S.D.
51	Read technical manuals, field manuals	35	4.80	0.53
1	Perform operator maintenance checks an	35	4.69	0.58
3	Inspect mechanical systems	35	4.46	1.22
5	Repair mechanical systems	35	4.46	1.22
6	Troubleshoot mechanical systems	35	4.43	1.27
19	Operate wheeled vehicles	35	4.31	0.87
11	Repair electrical systems	35	4.17	0.95
9	Inspect electrical systems	35	4.14	1.24
13	Troubleshoot electrical systems	35	4.09	1.29
75	Protect against NBC hazards	35	4.09	1.20
78	Fire individual weapons	35	3.97	1.10
88	Give first aid	35	3.80	1.28
82	Survive in the field	35	3.69	1.59
2	Perform operator checks and services o	35	3.63	1.57
97	Act as model	35	3.40	1.59
52	Use maps	35	3.14	1.44
81	Navigate	35	2.94	1.59
56	Give directions and instructions	35	2.91	1.38
80	Customs and laws of war	35	2.86	1.57
50	Use hand and arm signals	35	2.83	1.72
86	Place and camouflage tactical equipme	35	2.80	1.83
74	Use hand grenades	35	2.74	1.67
31	Operate gas and electric powered equi	35	2.60	1.68
54	Give oral reports	35	2.57	1.63
89	Detect and identify targets	35	2.54	1.72
77	Engage in hand-to-hand combat	35	2.40	1.79
47	Prepare technical forms and documents	35	2.14	2.05
53	Send and receive radio messages	35	2.11	1.81
23	Paint	35	2.09	1.56
15	Pack and load materials	35	2.06	1.63
92	Provide information	35	1.94	1.89
64	Order equipment and supplies	35	1.91	2.02
12	Repair electronic components	35	1.83	1.99
20	Operate track vehicles	35	1.77	1.96
48	Record, file, and dispatch informatio	35	1.54	1.80
26	Repair metal	35	1.51	1.54
14	Troubleshoot electronic components	35	1.37	2.03
79	Control individuals and crowds	35	1.37	1.61
42	Sketch maps, overlays, or range cards	35	1.34	1.70
22	Operate lifting, loading, and grading	35	1.31	1.81
76	Handle demolitions or mines	35	1.20	1.69
94	Support	35	1.17	1.67
90	Plan, organize, monitor	35	1.14	1.75
49	Receive, store, and issue supplies, e	35	1.06	1.63
8	Install electronic components	35	1.03	1.65
95	Train, develop	35	0.94	1.61
66	Plan placement or use of tactical equ	35	0.91	1.65
91	Clarify roles, provide feedback	35	0.77	1.50
24	Install wire and cables	35	0.71	1.15
93	Recognize, reward	35	0.71	1.32
10	Inspect electronic systems	35	0.69	1.55
35	Operate electronic equipment	34	0.65	1.10

Table 3b continued
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 63B: Vehicle/Generator Mechanic

No.	Task Category	N	MEAN	S.D.
7	Troubleshoot weapons	35	0.63	1.33
96	Discipline, punish	35	0.60	1.29
61	Decode data	35	0.54	1.34
40	Conduct land surveys	35	0.51	1.27
59	Interview	35	0.51	1.34
65	Estimate time and cost of maintenance	35	0.51	1.22
98	Conduct tactical operations	35	0.43	1.14
4	Repair weapons	35	0.40	1.19
37	Operate computer hardware	35	0.40	0.98
46	Type	35	0.37	0.88
33	Use audiovisual equipment	35	0.26	0.85
45	Draw illustrations	35	0.26	0.61
25	Repair plastic and fiberglass	35	0.20	0.47
71	Compute statistics or other mathemati	35	0.17	0.62
17	Prepare equipment and supplies for ai	35	0.14	0.49
41	Provide medical or dental treatment	35	0.14	0.85
70	Determine firing data for indirect fi	35	0.14	0.55
84	Fire heavy direct fire weapons (e.g.,	35	0.14	0.85
28	Install pipe assemblies	35	0.11	0.53
34	Reproduce printed material	35	0.11	0.32
18	Operate power excavating equipment	35	0.09	0.51
57	Write documents and correspondence	35	0.09	0.51
60	Provide counseling and other interper	35	0.09	0.51
63	Analyze weather conditions	35	0.09	0.37
29	Construct wooden buildings and other	35	0.06	0.24
72	Provide programming and data processi	35	0.06	0.34
44	Draw maps and overlays	35	0.03	0.17
55	Receive clients, patients, guests	35	0.03	0.17
16	Prepare parachutes	35	0.00	0.00
21	Operate boats	35	0.00	0.00
27	Assemble steel structures	35	0.00	0.00
30	Construct masonry buildings and struc	35	0.00	0.00
32	Select, layout and clean medical or d	35	0.00	0.00
36	Operate radar	35	0.00	0.00
38	Cook	35	0.00	0.00
39	Perform medical laboratory procedures	35	0.00	0.00
43	Produce technical drawings	35	0.00	0.00
58	Write and deliver presentations	35	0.00	0.00
62	Analyze electronic signals	35	0.00	0.00
67	Translate foreign languages	35	0.00	0.00
68	Analyze intelligence data	35	0.00	0.00
69	Control money	35	0.00	0.00
73	Control air traffic	35	0.00	0.00
83	Load and unload field artillery or ta	35	0.00	0.00
85	Prepare heavy weapons for tactical us	35	0.00	0.00
87	Fire indirect fire weapons (e.g., fie	35	0.00	0.00

Table 3c
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 71L: Administrative Specialist

No.	Task Category	N	MEAN	S.D.
46	Type	39	4.82	0.45
47	Prepare technical forms and documents	39	4.51	0.64
48	Record, file, and dispatch informatio	39	4.51	0.68
51	Read technical manuals, field manuals	39	3.77	1.44
37	Operate computer hardware	39	3.74	1.04
49	Receive, store, and issue supplies, e	39	3.36	1.50
56	Give directions and instructions	39	3.28	1.21
57	Write documents and correspondence	39	3.21	1.63
88	Give first aid	39	3.18	1.52
78	Fire individual weapons	39	3.15	1.73
75	Protect against NBC hazards	39	3.13	1.69
82	Survive in the field	39	3.03	1.72
34	Reproduce printed material	39	3.00	1.34
97	Act as model	39	2.92	1.87
2	Perform operator checks and services o	39	2.77	1.77
54	Give oral reports	39	2.56	1.45
52	Use maps	39	2.31	1.54
55	Receive clients, patients, guests	39	2.28	1.78
19	Operate wheeled vehicles	39	2.23	1.35
81	Navigate	39	2.21	1.64
80	Customs and laws of war	38	2.16	1.53
64	Order equipment and supplies	39	2.15	1.53
35	Operate electronic equipment	39	2.13	1.75
74	Use hand grenades	39	2.03	1.63
53	Send and receive radio messages	39	2.00	1.62
89	Detect and identify targets	39	1.95	1.56
1	Perform operator maintenance checks an	39	1.87	1.59
92	Provide information	38	1.87	1.71
77	Engage in hand-to-hand combat	39	1.72	1.57
94	Support	39	1.72	1.82
90	Plan, organize, monitor	39	1.62	1.80
8	Install electronic components	39	1.54	1.73
95	Train, develop	39	1.49	1.78
59	Interview	39	1.38	1.68
86	Place and camouflage tactical equipme	39	1.33	1.66
93	Recognize, reward	39	1.15	1.57
91	Clarify roles, provide feedback	38	1.13	1.63
31	Operate gas and electric powered equi	39	1.13	1.36
76	Handle demolitions or mines	39	1.13	1.70
42	Sketch maps, overlays, or range cards	39	1.08	1.42
50	Use hand and arm signals	39	1.08	1.33
96	Discipline, punish	39	1.03	1.55
58	Write and deliver presentations	39	0.90	1.29
79	Control individuals and crowds	39	0.90	1.23
72	Provide programming and data processi	39	0.87	1.49
69	Control money	39	0.82	1.65
7	Troubleshoot weapons	39	0.79	1.56
71	Compute statistics or other mathemati	39	0.79	1.26
15	Pack and load materials	39	0.77	1.13
60	Provide counseling and other interper	39	0.74	1.35
61	Decode data	39	0.74	1.37
40	Conduct land surveys	39	0.67	1.24

Table 3c continued
Task Category Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 71L: Administrative Specialist

No.	Task Category	N	MEAN	S.D.
45	Draw illustrations	39	0.64	1.09
33	Use audiovisual equipment	39	0.62	1.21
24	Install wire and cables	39	0.56	1.14
23	Paint	39	0.51	0.60
66	Plan placement or use of tactical equ	39	0.46	1.00
98	Conduct tactical operations	39	0.46	1.21
9	Inspect electrical systems	39	0.44	1.12
70	Determine firing data for indirect fi	39	0.36	1.09
3	Inspect mechanical systems	39	0.33	0.96
44	Draw maps and overlays	39	0.28	0.94
4	Repair weapons	39	0.21	0.73
68	Analyze intelligence data	39	0.21	0.80
87	Fire indirect fire weapons (e.g., fie	39	0.21	0.80
6	Troubleshoot mechanical systems	39	0.18	0.79
10	Inspect electronic systems	39	0.18	0.79
12	Repair electronic components	39	0.15	0.81

Table 4a

Job Activity Importance, Frequency, and Difficulty Rating Descriptive Statistics

11B Infantryman

No.	Job Activity	Importance N	Importance MEAN	Importance S.D.	Frequency N	Frequency MEAN	Frequency S.D.	Difficulty N	Difficulty MEAN	Difficulty S.D.
1	Work in a team	60	4.68	0.57	60	2.87	0.39	60	1.77	0.65
2	Lead a team	60	3.30	1.96	60	1.68	1.16	47	2.30	0.78
3	Support/counsel peers	60	2.65	1.86	60	1.35	1.01	44	1.84	0.71
4	Support/counsel subordinates	60	1.95	1.96	60	0.98	1.08	33	2.03	0.81
5	Coach peers	60	3.25	1.76	60	1.68	1.03	50	1.84	0.77
6	Coach subordinates	59	3.05	1.70	59	1.61	1.05	48	1.90	0.78
7	Make oral reports (to individuals)	60	3.28	1.53	60	1.85	0.95	53	1.75	0.62
8	Make oral reports (to groups)	60	2.10	1.92	60	1.02	1.00	37	1.92	0.68
9	Relay oral instructions	60	3.60	1.52	60	1.95	0.93	54	1.67	0.70
10	Ask questions	60	2.85	2.06	60	1.58	1.18	43	1.44	0.63
11	Record information	60	2.92	1.81	60	1.58	1.05	47	1.68	0.73
12	Write brief messages	60	2.57	1.83	60	1.47	1.11	44	1.48	0.70
13	Write longer reports	60	0.28	0.94	60	0.13	0.47	6	2.00	0.63
14	Monitor/interpret verbal messages	60	2.77	1.96	60	1.48	1.14	43	1.84	0.75
15	Recall verbal information	60	3.68	1.50	60	2.03	0.94	55	1.69	0.72
16	Monitor/interpret numerical informati	60	1.87	2.05	60	1.00	1.16	30	1.90	0.71
17	Recall numerical information	60	2.62	2.04	60	1.33	1.08	40	1.77	0.73
18	Monitor/interpret figural information	60	2.55	2.00	60	1.33	1.10	40	1.95	0.68
19	Recall figural information	60	2.72	1.90	60	1.42	1.08	43	1.98	0.67
20	Follow oral directions	59	4.59	0.62	59	2.73	0.52	59	1.75	0.63
21	Follow written directions	60	4.15	1.05	60	2.42	0.74	58	1.50	0.60
22	Judge size and distance	60	3.87	1.36	60	2.07	0.86	55	1.71	0.57
23	Judge location	60	4.00	1.30	60	2.07	0.84	56	2.16	0.63
24	Judge paths of moving objects	60	2.50	1.83	60	1.27	1.04	41	1.88	0.56
25	Solve electrical system problems	60	1.37	1.86	60	0.65	0.92	24	2.33	0.56
26	Solve mechanical system problems	60	2.20	1.93	60	1.08	1.01	37	2.14	0.71
27	Solve logistical problems	59	0.47	1.24	59	0.25	0.66	9	2.22	0.67
28	Solve tactical maneuver problems	60	1.88	2.16	60	0.97	1.16	28	2.21	0.74
29	Solve administrative problems	60	0.52	1.27	60	0.37	0.84	11	2.27	0.65
30	Solve leadership problems	60	1.60	2.05	60	0.73	0.99	25	2.52	0.65
31	Solve medical problems	60	0.73	1.66	60	0.37	0.82	11	2.27	0.65
32	Solve communication problems	60	1.37	1.92	60	0.65	0.95	22	1.95	0.79
33	Operate precision hand-held equipment	60	0.52	1.37	60	0.25	0.68	9	2.11	0.78
34	Operate hand-held tools	60	2.28	1.81	60	1.40	1.15	42	1.21	0.52
35	Operate hand-held power equipment	60	0.72	1.47	60	0.40	0.83	14	1.43	0.51
36	Operate larger power equipment	60	0.12	0.64	60	0.07	0.41	2	1.50	0.71
37	Operate full keyboard	60	0.42	1.06	60	0.22	0.58	8	2.00	0.53
38	Operate numeric keyboard	60	0.12	0.69	60	0.03	0.18	2	2.50	0.71
39	Adjust control device using one limb	60	2.10	1.98	60	1.27	1.21	34	1.38	0.60
40	Adj control device using multiple lim	60	1.98	2.02	60	1.12	1.17	32	1.53	0.62
41	Drive tracked vehicle	60	3.45	1.64	60	1.93	1.02	52	1.62	0.69
42	Drive heavy wheeled vehicle	60	2.30	1.96	60	1.22	1.15	38	1.63	0.75
43	Drive light wheeled vehicle	60	3.25	1.59	60	1.83	0.98	54	1.39	0.56
44	Aim stationary target	59	4.80	0.45	59	2.54	0.65	59	1.64	0.64
45	Aim moving target	59	4.69	0.77	59	2.47	0.73	58	2.10	0.69
46	Walk long distances	59	4.08	1.07	59	2.49	0.68	58	1.62	0.79
47	Run short distances	59	4.36	0.71	59	2.73	0.52	59	1.46	0.68
48	Push, pull, lift heavy weights	59	3.58	1.45	59	2.24	0.92	53	1.55	0.70
49	Throw objects	59	3.81	1.22	59	2.05	0.82	56	1.36	0.59
50	Sort, fold, feed by hand	59	1.59	1.81	59	0.85	1.00	29	1.17	0.38
51	Make coordinated movements	59	3.39	1.61	59	1.80	0.98	52	1.54	0.61
52	Work long hours	59	3.93	1.27	59	2.41	0.81	57	1.84	0.75
53	Work under adverse conditions	59	4.15	1.03	59	2.41	0.75	58	1.86	0.74

Table 4b
Job Activity Importance, Frequency, and Difficulty Rating Descriptive Statistics
63B Vehicle/Generator Mechanic

No.	Job Activity	Importance N	Importance MEAN	Importance S.D.	Frequency N	Frequency MEAN	Frequency S.D.	Difficulty N	Difficulty MEAN	Difficulty S.D.
1	Work in a team	35	3.66	1.41	35	2.20	0.96	33	1.73	0.76
2	Lead a team	35	1.77	1.93	35	0.86	1.03	18	2.11	0.83
3	Support/counsel peers	35	2.06	1.92	35	1.17	1.15	21	1.95	0.80
4	Support/counsel subordinates	35	1.69	1.91	35	0.91	1.07	18	2.28	0.75
5	Coach peers	34	2.32	1.74	34	1.29	1.06	24	1.92	0.65
6	Coach subordinates	34	2.41	1.88	34	1.24	1.05	24	2.00	0.72
7	Make oral reports (to individuals)	35	3.00	1.81	35	1.71	1.15	29	1.83	0.80
8	Make oral reports (to groups)	35	0.54	0.95	35	0.37	0.60	11	2.09	0.83
9	Relay oral instructions	35	2.89	1.59	35	1.60	0.98	31	1.61	0.62
10	Ask questions	35	2.57	2.20	35	1.57	1.31	23	1.35	0.57
11	Record information	34	1.32	1.90	34	0.71	1.06	12	1.67	0.65
12	Write brief messages	35	1.09	1.60	35	0.66	0.97	13	1.69	0.75
13	Write longer reports	35	0.14	0.43	35	0.11	0.32	4	2.75	0.50
14	Monitor/interpret verbal messages	35	1.86	1.87	35	1.03	1.07	20	1.75	0.72
15	Recall verbal information	35	3.03	1.72	35	1.66	1.03	29	1.83	0.71
16	Monitor/interpret numerical informati	34	1.29	1.77	34	0.76	1.02	14	1.79	0.70
17	Recall numerical information	35	2.51	1.74	35	1.43	0.98	28	1.96	0.74
18	Monitor/interpret figural information	35	3.11	1.78	35	1.71	1.02	31	1.97	0.60
19	Recall figural information	34	2.50	1.96	34	1.29	1.03	25	1.96	0.73
20	Follow oral directions	35	4.49	0.92	35	2.63	0.60	35	1.77	0.69
21	Follow written directions	35	4.60	0.81	35	2.71	0.52	35	1.89	0.63
22	Judge size and distance	35	2.71	1.72	35	1.57	1.09	28	1.61	0.79
23	Judge location	35	2.83	1.60	35	1.37	1.00	30	1.97	0.81
24	Judge paths of moving objects	35	2.83	1.52	35	1.60	0.98	32	1.69	0.82
25	Solve electrical system problems	35	4.20	1.28	35	2.26	0.82	33	2.55	0.56
26	Solve mechanical system problems	35	4.34	1.30	35	2.43	0.78	33	2.36	0.65
27	Solve logistical problems	35	1.54	1.82	35	0.77	1.00	17	2.18	0.81
28	Solve tactical maneuver problems	35	0.46	1.22	35	0.14	0.36	5	2.40	0.89
29	Solve administrative problems	35	0.91	1.38	35	0.37	0.55	12	2.17	0.72
30	Solve leadership problems	35	0.69	1.25	35	0.29	0.52	9	2.00	0.71
31	Solve medical problems	35	0.31	0.99	35	0.11	0.32	4	2.00	0.00
32	Solve communication problems	35	0.77	1.50	35	0.37	0.73	9	2.44	0.73
33	Operate precision hand-held equipment	35	3.69	1.51	35	1.89	0.96	32	2.06	0.67
34	Operate hand-held tools	35	4.60	0.88	35	2.74	0.56	35	1.46	0.70
35	Operate hand-held power equipment	35	3.94	1.45	35	2.20	0.93	33	1.61	0.70
36	Operate larger power equipment	35	1.54	1.93	35	0.80	1.08	15	1.60	0.63
37	Operate full keyboard	35	0.29	0.75	35	0.17	0.45	5	1.40	0.55
38	Operate numeric keyboard	35	0.31	0.96	35	0.11	0.32	4	1.75	0.50
39	Adjust control device using one limb	35	3.00	1.88	35	1.74	1.09	28	1.54	0.74
40	Adj control device using multiple lim	35	3.49	1.70	35	2.06	1.03	31	1.71	0.78
41	Drive tracked vehicle	35	1.74	1.90	35	0.94	1.11	18	1.83	0.51
42	Drive heavy wheeled vehicle	35	2.80	2.01	35	1.60	1.22	25	1.72	0.68
43	Drive light wheeled vehicle	35	4.34	0.94	35	2.54	0.61	35	1.57	0.70
44	Aim stationary target	35	3.11	1.41	34	1.35	0.81	32	1.53	0.62
45	Aim moving target	35	2.14	1.72	34	0.88	0.77	24	1.63	0.65
46	Walk long distances	35	2.71	1.64	34	1.32	0.91	29	1.24	0.58
47	Run short distances	35	2.69	1.66	34	1.82	1.11	30	1.33	0.61
48	Push, pull, lift heavy weights	35	3.60	1.19	33	2.15	0.83	33	1.64	0.65
49	Throw objects	35	1.46	1.74	35	0.83	0.98	18	1.11	0.47
50	Sort, fold, feed by hand	35	1.37	1.61	35	0.74	0.85	19	1.16	0.50
51	Make coordinated movements	35	3.31	1.53	35	1.77	1.00	32	1.50	0.62
52	Work long hours	35	3.80	1.28	35	2.09	0.89	34	1.76	0.78
53	Work under adverse conditions	35	3.80	1.32	35	2.06	0.87	34	1.94	0.69

Table 4c
Job Activity Importance, Frequency, and Difficulty Rating Descriptive Statistics
71L Administrative Specialist

		Importance			Frequency			Difficulty		
No.	Job Activity	N	MEAN	S.D.	N	MEAN	S.D.	N	MEAN	S.D.
1	Work in a team	39	3.72	1.45	39	2.10	1.02	36	1.53	0.74
2	Lead a team	39	1.59	1.90	39	0.79	1.03	19	2.11	0.81
3	Support/counsel peers	39	2.64	1.58	39	1.23	0.81	31	1.84	0.82
4	Support/counsel subordinates	39	1.85	1.97	39	0.82	0.91	20	2.00	0.79
5	Coach peers	39	3.36	1.39	39	1.56	0.68	36	1.97	0.70
6	Coach subordinates	39	2.79	1.88	39	1.31	0.92	29	1.79	0.73
7	Make oral reports (to individuals)	39	3.00	1.49	39	1.72	1.00	34	1.68	0.64
8	Make oral reports (to groups)	39	1.13	1.47	39	0.59	0.79	17	2.18	0.73
9	Relay oral instructions	39	3.28	1.52	39	1.85	0.96	34	1.74	0.62
10	Ask questions	39	2.00	1.85	39	1.21	1.17	23	1.61	0.66
11	Record information	39	4.03	1.04	39	2.62	0.63	38	1.55	0.69
12	Write brief messages	39	2.87	1.98	39	1.72	1.26	28	1.75	0.80
13	Write longer reports	39	0.72	1.38	39	0.36	0.74	9	1.89	0.78
14	Monitor/interpret verbal messages	39	2.95	1.64	39	1.77	1.04	32	1.78	0.66
15	Recall verbal information	39	2.92	1.72	39	1.74	1.07	31	1.71	0.59
16	Monitor/interpret numerical informati	39	1.69	1.73	39	0.92	1.04	21	1.67	0.66
17	Recall numerical information	39	1.95	1.81	39	1.13	1.13	23	1.83	0.72
18	Monitor/interpret figural information	39	1.26	1.43	39	0.69	0.83	19	1.84	0.60
19	Recall figural information	39	1.13	1.28	39	0.64	0.78	19	1.89	0.74
20	Follow oral directions	39	4.21	0.89	39	2.56	0.60	38	1.71	0.52
21	Follow written directions	39	4.36	0.58	39	2.62	0.54	38	1.76	0.63
22	Judge size and distance	39	0.97	1.31	39	0.49	0.64	16	1.69	0.70
23	Judge location	39	1.62	1.57	39	0.72	0.60	25	1.76	0.78
24	Judge paths of moving objects	39	0.51	0.97	39	0.26	0.44	10	1.70	0.82
25	Solve electrical system problems	39	0.49	1.07	39	0.23	0.48	8	2.25	0.71
26	Solve mechanical system problems	39	0.54	1.05	39	0.26	0.44	10	2.20	0.63
27	Solve logistical problems	38	0.50	1.03	38	0.26	0.55	8	1.63	0.52
28	Solve tactical maneuver problems	39	0.21	0.66	39	0.13	0.41	4	2.00	0.82
29	Solve administrative problems	39	3.13	1.54	39	1.85	0.96	34	1.97	0.72
30	Solve leadership problems	39	1.36	1.66	39	0.69	0.86	18	2.17	0.79
31	Solve medical problems	39	0.10	0.45	39	0.05	0.22	2	2.00	1.41
32	Solve communication problems	39	0.95	1.45	39	0.51	0.82	13	1.85	0.80
33	Operate precision hand-held equipment	39	0.26	0.91	39	0.13	0.52	3	2.33	0.58
34	Operate hand-held tools	39	0.33	0.81	39	0.26	0.59	8	1.38	0.74
35	Operate hand-held power equipment	39	0.13	0.66	39	0.10	0.50	2	2.00	0.00
36	Operate larger power equipment	39	0.13	0.66	39	0.10	0.50	2	2.50	0.71
37	Operate full keyboard	39	4.69	0.57	39	2.95	0.22	39	1.87	0.57
38	Operate numeric keyboard	39	3.00	2.08	39	1.79	1.32	28	1.79	0.69
39	Adjust control device using one limb	39	0.72	1.23	39	0.49	0.82	12	1.25	0.45
40	Adj control device using multiple lim	39	0.41	0.94	39	0.28	0.65	7	1.29	0.49
41	Drive tracked vehicle	39	0.33	1.11	39	0.18	0.60	4	2.25	0.50
42	Drive heavy wheeled vehicle	39	0.64	1.39	39	0.31	0.69	8	1.88	0.64
43	Drive light wheeled vehicle	39	2.13	1.51	39	1.15	0.78	32	1.34	0.48
44	Aim stationary target	39	2.13	1.87	39	0.92	0.84	26	1.65	0.56
45	Aim moving target	39	1.13	1.79	39	0.46	0.68	14	2.14	0.66
46	Walk long distances	39	1.41	1.60	39	0.69	0.73	22	1.55	0.67
47	Run short distances	39	2.10	1.62	39	1.33	1.03	29	1.28	0.45
48	Push, pull, lift heavy weights	39	1.10	1.31	39	0.64	0.71	20	1.40	0.50
49	Throw objects	39	0.92	1.38	39	0.49	0.60	17	1.18	0.39
50	Sort, fold, feed by hand	39	2.72	1.34	39	1.64	0.90	35	1.11	0.32
51	Make coordinated movements	39	1.54	1.50	39	0.92	0.90	23	1.30	0.47
52	Work long hours	39	2.49	1.62	39	1.23	0.90	30	1.43	0.57
53	Work under adverse conditions	39	1.92	1.74	39	0.79	0.80	24	1.67	0.70

Table 5a
Job Activity Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 11B: Infantryman

No.	Job Activity	N	MEAN	S.D.
44	Aim stationary target	59	4.80	0.45
45	Aim moving target	59	4.69	0.77
1	Work in a team	60	4.68	0.57
20	Follow oral directions	59	4.59	0.62
47	Run short distances	59	4.36	0.71
53	Work under adverse conditions	59	4.15	1.03
21	Follow written directions	60	4.15	1.05
46	Walk long distances	59	4.08	1.07
23	Judge location	60	4.00	1.30
52	Work long hours	59	3.93	1.27
22	Judge size and distance	60	3.87	1.36
49	Throw objects	59	3.81	1.22
15	Recall verbal information	60	3.68	1.50
9	Relay oral instructions	60	3.60	1.52
48	Push, pull, lift heavy weights	59	3.58	1.45
41	Drive tracked vehicle	60	3.45	1.64
51	Make coordinated movements	59	3.39	1.61
2	Lead a team	60	3.30	1.96
7	Make oral reports (to individuals)	60	3.28	1.53
5	Coach peers	60	3.25	1.76
43	Drive light wheeled vehicle	60	3.25	1.59
6	Coach subordinates	59	3.05	1.70
11	Record information	60	2.92	1.81
10	Ask questions	60	2.85	2.06
14	Monitor/interpret verbal messages	60	2.77	1.96
19	Recall figural information	60	2.72	1.90
3	Support/counsel peers	60	2.65	1.86
17	Recall numerical information	60	2.62	2.04
12	Write brief messages	60	2.57	1.83
18	Monitor/interpret figural information	60	2.55	2.00
24	Judge paths of moving objects	60	2.50	1.83
42	Drive heavy wheeled vehicle	60	2.30	1.96
34	Operate hand-held tools	60	2.28	1.81
26	Solve mechanical system problems	60	2.20	1.93
8	Make oral reports (to groups)	60	2.10	1.92
39	Adjust control device using one limb	60	2.10	1.98
40	Adj control device using multiple lim	60	1.98	2.02
4	Support/counsel subordinates	60	1.95	1.96
28	Solve tactical maneuver problems	60	1.88	2.16
16	Monitor/interpret numerical informati	60	1.87	2.05
30	Solve leadership problems	60	1.60	2.05
50	Sort, fold, feed by hand	59	1.59	1.81
25	Solve electrical system problems	60	1.37	1.86
32	Solve communication problems	60	1.37	1.92
31	Solve medical problems	60	0.73	1.66
35	Operate hand-held power equipment	60	0.72	1.47
29	Solve administrative problems	60	0.52	1.27
33	Operate precision hand-held equipment	60	0.52	1.37
27	Solve logistical problems	59	0.47	1.24
37	Operate full keyboard	60	0.42	1.06
13	Write longer reports	60	0.28	0.94
36	Operate larger power equipment	60	0.12	0.64
38	Operate numeric keyboard	60	0.12	0.69

Table 5b
Job Activity Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 63B: Vehicle/ Generator Mechanic

No.	Job Activity	N	MEAN	S.D.
21	Follow written directions	35	4.60	0.81
34	Operate hand-held tools	35	4.60	0.88
20	Follow oral directions	35	4.49	0.92
26	Solve mechanical system problems	35	4.34	1.30
43	Drive light wheeled vehicle	35	4.34	0.94
25	Solve electrical system problems	35	4.20	1.28
35	Operate hand-held power equipment	35	3.94	1.45
52	Work long hours	35	3.80	1.28
53	Work under adverse conditions	35	3.80	1.32
33	Operate precision hand-held equipment	35	3.69	1.51
1	Work in a team	35	3.66	1.41
48	Push, pull, lift heavy weights	35	3.60	1.19
40	Adj control device using multiple lim	35	3.49	1.70
51	Make coordinated movements	35	3.31	1.53
18	Monitor/interpret figural information	35	3.11	1.78
44	Aim stationary target	35	3.11	1.41
15	Recall verbal information	35	3.03	1.72
7	Make oral reports (to individuals)	35	3.00	1.81
39	Adjust control device using one limb	35	3.00	1.88
9	Relay oral instructions	35	2.89	1.59
23	Judge location	35	2.83	1.60
24	Judge paths of moving objects	35	2.83	1.52
42	Drive heavy wheeled vehicle	35	2.80	2.01
22	Judge size and distance	35	2.71	1.72
46	Walk long distances	35	2.71	1.64
47	Run short distances	35	2.69	1.66
10	Ask questions	35	2.57	2.20
17	Recall numerical information	35	2.51	1.74
19	Recall figural information	34	2.50	1.96
6	Coach subordinates	34	2.41	1.88
5	Coach peers	34	2.32	1.74
45	Aim moving target	35	2.14	1.72
3	Support/counsel peers	35	2.06	1.92
14	Monitor/interpret verbal messages	35	1.86	1.87
2	Lead a team	35	1.77	1.93
41	Drive tracked vehicle	35	1.74	1.90
4	Support/counsel subordinates	35	1.69	1.91
27	Solve logistical problems	35	1.54	1.82
36	Operate larger power equipment	35	1.54	1.93
49	Throw objects	35	1.46	1.74
50	Sort, fold, feed by hand	35	1.37	1.61
11	Record information	34	1.32	1.90
16	Monitor/interpret numerical informati	34	1.29	1.77
12	Write brief messages	35	1.09	1.60
29	Solve administrative problems	35	0.91	1.38
32	Solve communication problems	35	0.77	1.50
30	Solve leadership problems	35	0.69	1.25
8	Make oral reports (to groups)	35	0.54	0.95
28	Solve tactical maneuver problems	35	0.46	1.22
31	Solve medical problems	35	0.31	0.99
38	Operate numeric keyboard	35	0.31	0.96
37	Operate full keyboard	35	0.29	0.75
13	Write longer reports	35	0.14	0.43

Table 5c
Job Activity Importance Rating Descriptive Statistics
Sorted by Mean Rating
For 71L: Administrative Specialist

No.	Job Activity	N	MEAN	S.D.
37	Operate full keyboard	39	4.69	0.57
21	Follow written directions	39	4.36	0.58
20	Follow oral directions	39	4.21	0.89
11	Record information	39	4.03	1.04
1	Work in a team	39	3.72	1.45
5	Coach peers	39	3.36	1.39
9	Relay oral instructions	39	3.28	1.52
29	Solve administrative problems	39	3.13	1.54
7	Make oral reports (to individuals)	39	3.00	1.49
38	Operate numeric keyboard	39	3.00	2.08
14	Monitor/interpret verbal messages	39	2.95	1.64
15	Recall verbal information	39	2.92	1.72
12	Write brief messages	39	2.87	1.98
6	Coach subordinates	39	2.79	1.88
50	Sort, fold, feed by hand	39	2.72	1.34
3	Support/counsel peers	39	2.64	1.58
52	Work long hours	39	2.49	1.62
43	Drive light wheeled vehicle	39	2.13	1.51
44	Aim stationary target	39	2.13	1.87
47	Run short distances	39	2.10	1.62
10	Ask questions	39	2.00	1.85
17	Recall numerical information	39	1.95	1.81
53	Work under adverse conditions	39	1.92	1.74
4	Support/counsel subordinates	39	1.85	1.97
16	Monitor/interpret numerical informati	39	1.69	1.73
23	Judge location	39	1.62	1.57
2	Lead a team	39	1.59	1.90
51	Make coordinated movements	39	1.54	1.50
46	Walk long distances	39	1.41	1.60
30	Solve leadership problems	39	1.36	1.66
18	Monitor/interpret figural information	39	1.26	1.43
8	Make oral reports (to groups)	39	1.13	1.47
19	Recall figural information	39	1.13	1.28
45	Aim moving target	39	1.13	1.79
48	Push, pull, lift heavy weights	39	1.10	1.31
22	Judge size and distance	39	0.97	1.31
32	Solve communication problems	39	0.95	1.45
49	Throw objects	39	0.92	1.38
13	Write longer reports	39	0.72	1.38
39	Adjust control device using one limb	39	0.72	1.23
42	Drive heavy wheeled vehicle	39	0.64	1.39
26	Solve mechanical system problems	39	0.54	1.05
24	Judge paths of moving objects	39	0.51	0.97
27	Solve logistical problems	38	0.50	1.03
25	Solve electrical system problems	39	0.49	1.07
40	Adj control device using multiple lim	39	0.41	0.94
34	Operate hand-held tools	39	0.33	0.81
41	Drive tracked vehicle	39	0.33	1.11
33	Operate precision hand-held equipment	39	0.26	0.91
28	Solve tactical maneuver problems	39	0.21	0.66
35	Operate hand-held power equipment	39	0.13	0.66
36	Operate larger power equipment	39	0.13	0.66
31	Solve medical problems	39	0.10	0.45

Table 6a
Task Category Mean Within-SME Correlations*
between Importance, Frequency, and Difficulty Ratings

		Importance	Frequency
11B	Difficulty	.11	-.09
(n=60)	Frequency	.48	
63B	Difficulty	.25	.13
(n=35)	Frequency	.57	
71L	Difficulty	.22	.14
(n=39)	Frequency	.63	

* Ratings set to missing when not part of job

Table 6b
Job Activity Mean Within-SME Correlations*
between Importance, Frequency, and Difficulty Ratings

		Importance	Frequency
11B	Difficulty	.08	-.02
(n=60)	Frequency	.56	
63B	Difficulty	.21	.14
(n=35)	Frequency	.70	
71L	Difficulty	.20	.07
(n=39)	Frequency	.70	

* Ratings set to missing when not part of job

Table 7

Variance Component and Reliability Estimates
For Task Category and Activity Importance Ratings
Across 3 MOS

VARIANCE COMPONENTS

	Task		Activity	
Components	Variance	Percent	Variance	Percent
Component	1.059	30.51%	0.495	13.48%
MOS	0.025	0.73%	0.111	3.02%
Component x MOS	0.921	26.54%	1.062	28.91%
Rank	0.000	0.00%	0.000	0.00%
Command	0.000	0.00%	0.002	0.06%
Rank x Command	0.041	1.19%	0.000	0.00%
Rank x MOS	0.036	1.04%	0.108	2.94%
Command x MOS	0.045	1.30%	0.032	0.88%
Rk. x C x M	0.003	0.09%	0.000	0.00%
Component x Rank	0.038	1.09%	0.051	1.39%
Component x Command	0.039	1.12%	0.035	0.95%
Comp x Rank x Command	0.000	0.00%	0.000	0.00%
Comp x Rank x MOS	0.027	0.78%	0.033	0.91%
Comp x Command x MOS	0.068	1.96%	0.064	1.74%
Rank x Command x MOS	0.003	0.09%	0.000	0.00%
Cp x Rk x Cd x M	0.101	2.91%	0.000	0.00%
Rater(Rk Cd M)	0.096	2.77%	0.184	5.02%
Component x Rater	0.968	27.88%	1.495	40.71%
Total	3.470	100.00%	3.673	100.00%

RELIABILITY

	Task	Activity
Overall	0.58	0.45
Within Rank	0.62	0.48
Within Command	0.63	0.47
Within Rank & Command	0.65	0.50

Table 8a

Variance Component and Reliability Estimates
For Task Category and Activity Importance Ratings
NCOs Across 3 MOS

	Task Category Variance	Task Category Percent	Job Activity Variance	Job Activity Percent
VARIANCE COMPONENTS				
Component	1.03	29.89%	0.57	15.14%
MOS	0.07	2.13%	0.24	6.23%
Component*MOS	1.01	29.22%	1.05	27.69%
Command	0.04	1.27%	0.00	0.00%
Component*Command	0.00	0.00%	0.00	0.00%
MOS*Command	0.06	1.86%	0.02	0.53%
Rater(Cd M)	0.11	3.29%	0.18	4.80%
Component*Rater(Cd M)	1.11	32.33%	1.73	45.60%
	3.44	100.00%	3.79	100.00%
RELIABILITY				
Overall	0.61		0.49	
Within Command	0.63		0.49	

Table 8b

Variance Component and Reliability Estimates
For Task Category and Activity Importance Ratings
Officers Across 3 MOS

	Task Category Variance	Task Category Percent	Job Activity Variance	Job Activity Percent
VARIANCE COMPONENTS				
Component	1.31	39.16%	0.46	14.15%
MOS	0.04	1.12%	0.17	5.30%
Component*MOS	0.77	22.95%	1.24	38.19%
Command	0.00	0.00%	0.00	0.00%
Component*Command	0.00	0.00%	0.05	1.66%
MOS*Command	0.02	0.54%	0.00	0.00%
Rater(Cd M)	0.13	3.74%	0.07	2.27%
Component*Rater(Cd M)	1.09	32.50%	1.25	38.43%
	3.35	100.00%	3.24	100.00%
RELIABILITY				
Overall	0.63		0.58	
Within Command	0.64		0.59	

Table 9a

Variance Component and Reliability Estimates
For Task Category and Activity Importance Ratings
FORSCOM SMEs Across 3 MOS

	VARIANCE COMPONENTS			
	Task Category		Job Activity	
	Variance	Percent	Variance	
Component	1.05	30.00%	0.50	14.53%
MOS	0.14	3.91%	0.27	7.82%
Component*MOS	1.09	31.28%	1.04	30.04%
Rank	0.00	0.00%	0.00	0.00%
Component*Rank	0.00	0.00%	0.01	0.27%
MOS*Rank	0.05	1.33%	0.05	1.42%
Rater(Rk M)	0.11	3.28%	0.13	3.77%
Component*Rater(Rk M)	1.06	30.20%	1.46	42.16%
	3.50	100.00%	3.45	100.00%

	RELIABILITY	
Overall	0.65	0.52
Within Command	0.66	0.53

Table 9b

Variance Component and Reliability Estimates
For Task Category and Activity Importance Ratings
DOTD SMEs Across 3 MOS

	VARIANCE COMPONENTS			
	Task Category		Job Activity	
	Variance	Percent	Variance	Percent
Component	1.14	35.30%	0.55	14.26%
MOS	.00	0.07%	0.03	0.88%
Component*MOS	0.89	27.76%	1.22	31.40%
Rank	0.02	0.58%	0.00	0.00%
Component*Rank	0.00	0.00%	0.00	0.00%
MOS*Rank	.00	0.05%	0.09	2.30%
Rater(Rk M)	0.15	4.65%	0.35	8.91%
Component*Rater(Rk M)	1.02	31.59%	1.64	42.25%
	3.22	100.00%	3.88	100.00%

	RELIABILITY	
Overall	0.63	0.47
Within Command	0.64	0.48

Table 10a

Variance Component and Reliability Estimates
For Task Category Importance Ratings
Within each MOS

VARIANCE COMPONENTS

Components	11B Variance	11B Percent	63B Variance	63B Percent	71L Variance	71L Percent
Task	2.21	50.76%	2.11	58.59%	1.48	50.60%
Rank	0.01	0.32%	0.01	0.39%	0.02	0.57%
Command	0.08	1.82%	0.01	0.37%	.00	0.10%
Rank x Command	0.08	1.86%	.00	0.04%	0.01	0.50%
Rater	0.12	2.76%	0.06	1.68%	0.16	5.38%
Task x Rater	1.67	38.33%	1.01	27.92%	1.21	41.44%
Task x Rank	0.10	2.37%	0.05	1.43%	0.01	0.47%
Task x Command	0.08	1.78%	0.24	6.68%	0.03	0.94%
Task x Rank x Command	0.00	0.00%	0.10	2.90%	0.00	0.00%
Total	4.35	100.00%	3.60	100.00%	2.93	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.51	0.59	0.51
Within Rank	0.52	0.61	0.51
Within Command	0.53	0.65	0.51
Within Rank & Command	0.55	0.66	0.52

Table 10b

Variance Component and Reliability Estimates
For Activity Importance Ratings
Within each MOS

VARIANCE COMPONENTS

Components	11B Variance	11B Percent	63B Variance	63B Percent	71L Variance	71L Percent
Activity	1.72	40.31%	1.40	36.47%	1.51	42.73%
Rank	0.03	0.62%	0.03	0.66%	0.03	0.91%
Command	0.03	0.74%	.00	0.06%	0.01	0.30%
Rank x Command	0.01	0.25%	0.01	0.30%	0.00	0.00%
Rater	0.35	8.16%	0.21	5.44%	0.23	6.62%
Activity x Rater	1.97	46.18%	1.56	40.59%	1.70	47.82%
Activity x Rank	0.11	2.67%	0.20	5.14%	0.00	0.00%
Activity x Command	0.05	1.07%	0.44	11.33%	0.00	0.00%
Act. x Rank x Command	0.00	0.00%	0.00	0.00%	0.06	1.61%
Total	4.27	100.00%	3.85	100.00%	3.55	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.40	0.36	0.43
Within Rank	0.42	0.39	0.44
Within Command	0.41	0.41	0.44
Within Rank & Command	0.43	0.44	0.44

Table 11a

Variance Component and Reliability Estimates
For Task Category Importance Ratings
For NCOs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Task	2.30	48.93%	2.08	57.15%	1.32	47.28%
Command	0.14	2.90%	0.01	0.37%	0.00	0.00%
Rater(Command)	0.19	4.06%	0.10	2.88%	0.19	6.69%
Task*Rater	1.98	41.95%	0.99	27.19%	1.29	46.03%
Task*Command	0.10	2.16%	0.45	12.41%	0.00	0.00%
	4.71	100.00%	3.64	100.00%	2.80	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.49	0.57	0.47
Within Command	0.52	0.66	0.47

Table 11b

Variance Component and Reliability Estimates
For Task Category Importance Ratings
For Officers within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Task	2.41	57.20%	2.28	62.94%	1.76	55.74%
Command	.00	0.02%	.00	0.03%	0.03	0.94%
Rater(Command)	0.20	4.83%	0.04	1.00%	0.16	5.14%
Task*Rater	1.60	37.95%	1.30	36.02%	1.20	38.19%
Task*Command	0.00	0.00%	0.00	0.00%	0.00	0.00%
	4.21	100.00%	3.62	100.00%	3.15	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.57	0.63	0.56
Within Command	0.57	0.63	0.56

Table 11c

Variance Component and Reliability Estimates
For Job Activity Importance Ratings
For NCOs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Act	1.76	37.90%	1.61	39.01%	1.33	38.77%
Command	0.03	0.56%	0.00	0.00%	0.03	0.81%
Rater(Command)	0.44	9.54%	0.30	7.26%	0.26	7.63%
Act*Rater	2.41	52.00%	1.82	44.24%	1.81	52.80%
Act*Command	0.00	0.00%	0.39	9.48%	0.00	0.00%
	4.63	100.00%	4.12	100.00%	3.43	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.38	0.39	0.39
Within Command	0.38	0.43	0.39

Table 11d

Variance Component and Reliability Estimates
For Job Activity Importance Ratings
For Officers within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Act	1.85	48.87%	1.39	40.63%	1.67	45.71%
Command	0.01	0.14%	0.00	0.00%	0.01	0.15%
Rater(Command)	0.23	6.01%	0.08	2.35%	0.26	7.05%
Act*Rater	1.70	44.99%	1.84	53.85%	1.64	45.03%
Act*Command	0.00	0.00%	0.11	3.16%	0.08	2.06%
	3.79	100.00%	3.41	100.00%	3.65	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.49	0.41	0.46
Within Command	0.49	0.42	0.47

Table 12a

Variance Component and Reliability Estimates
For Task Category Importance Ratings
For FORSCOM SMEs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Task	2.38	50.50%	1.91	55.64%	1.55	55.99%
Rank	0.03	0.69%	0.00	0.00%	.00	0.16%
Rater(Rank)	0.25	5.33%	0.11	3.33%	0.16	5.75%
Task*Rater	1.98	42.04%	1.40	41.03%	1.05	38.10%
Task*Rank	0.07	1.44%	0.00	0.00%	0.00	0.00%
	4.72	100.00%	3.42	100.00%	2.76	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.51	0.56	0.56
Within Rank	0.52	0.56	0.56

Table 12b

Variance Component and Reliability Estimates
For Task Category Importance Ratings
For DOTD SMEs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Task	2.13	54.79%	2.45	64.36%	1.51	48.41%
Rank	0.02	0.46%	0.02	0.39%	0.04	1.43%
Rater(Rank)	0.21	5.34%	0.05	1.33%	0.18	5.70%
Task*Rater	1.53	39.41%	1.22	32.08%	1.39	44.46%
Task*Rank	0.00	0.00%	0.07	1.83%	0.00	0.00%
	3.89	100.00%	3.80	100.00%	3.13	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.55	0.64	0.48
Within Rank	0.55	0.66	0.49

Table 12c

Variance Component and Reliability Estimates
For Job Activity Importance Ratings
For FORSCOM SMEs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Act	1.66	39.55%	1.20	34.98%	1.46	43.82%
Rank	0.01	0.25%	.00	0.01%	0.05	1.60%
Rater(Rank)	0.34	8.02%	0.41	12.01%	0.16	4.89%
Act*Rater	2.10	50.02%	1.82	52.75%	1.56	46.89%
Act*Rank	0.09	2.17%	0.01	0.26%	0.09	2.80%
	4.20	100.00%	3.44	100.00%	3.33	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.40	0.35	0.44
Within Rank	0.41	0.35	0.46

Table 12d

Variance Component and Reliability Estimates
For Job Activity Importance Ratings
For DOTD SMEs within each MOS

VARIANCE COMPONENTS

	11B		63B		71L	
	Variance	Percent	Variance	Percent	Variance	Percent
Act	1.81	41.74%	1.97	45.43%	1.52	41.04%
Rank	0.06	1.42%	0.12	2.70%	0.03	0.69%
Rater(Rank)	0.43	9.92%	0.03	0.73%	0.31	8.33%
Act*Rater	2.03	46.92%	2.13	49.14%	1.84	49.94%
Act*Rank	0.00	0.00%	0.09	2.00%	0.00	0.00%
	4.33	100.00%	4.34	100.00%	3.69	100.00%

RELIABILITY

	11B	63B	71L
Overall	0.42	0.45	0.41
Within Rank	0.42	0.48	0.41

Table 13a

Task Category and Job Activity
Importance Ratings Single-SME Reliability Estimates*
Across 3 MOS

	Task Category		Job Activity	
	Reliability	No. SMEs required for r_{xx}=90	Reliability	No. SMEs required for r_{xx}=90
Overall	58	7	45	11
NCO's	61	6	49	9
Officers	63	5	58	7
FORSCOM	65	5	52	8
DOTD	63	5	47	10

* Decimal point omitted

Table 13b

Task Category
Importance Ratings Single-SME Reliability Estimates*
Within each MOS

	11B				63B				71L			
	Pilot Test		Pre-Test		Pilot Test		Pre-Test		Pilot Test		Pre-Test	
	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90
Overall	51	9	63	5	59	6	53	8	51	9	44	11
NCO's	49	9	64	5	57	7	07	120	47	10	37	15
Officers	57	7	63	5	63	5	56	7	56	7	34	17
FORSCOM	51	9			56	7			56	7		
DOTD	55	7			64	5			48	10		

Table 13c

Job Activitiy
Importance Ratings Single-SME Reliability Estimates*
Within each MOS

	11B				63B				71L			
	Pilot Test		Pre-Test		Pilot Test		Pre-Test		Pilot Test		Pre-Test	
	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90	r_{xx}	N req'd r_{xx}=90
Overall	40	14	46	11	36	16	41	13	43	12	35	17
NCOs	38	15	54	8	39	14	36	16	39	14	37	15
Officers	49	9	40	14	41	13	56	7	46	11	34	17
FORSCOM	40	14			35	17			44	11		
DOTD	42	12			45	11			41	13		

* Decimal point omitted

Table 14a: Tasks Matched to Job Component with 50% or Greater Agreement

11B: Infantryman
N=60

Project A Tasks	Task Category High	%	2nd	%	3rd	%	Job Activity High	%	2nd	%	3rd	%
1: Prepare a dragon for firing	2	50%										
2: Tech movement in urban terrain							1	57%				
3: Call for/adjust indirect fire												
4: Select fire overwatch position												
5: Set headspace on .50 machinegun												
6: Place AN/PVS-5 into operation												
7: Zero AN/PVS-4 to an M16A1	78	58%										
8: Select Hasty Firing Position												
9: Operate radio set	53	60%										
10: Collect/report information	54	52%										
11: Recognize & identify vehicles	89	67%										
13: Indentify terrain features	52	75%										
14: Perform maintenance on M16A1	2	70%										
15: Load, reduce, clear M60												
16: Prepare range card for M60	42	63%										
17: Engage target with LAW												
18: Engage targets w/ hand grenades	74	73%					49	65%				
19: Install & Fire Claymore mine	76	55%										
20: Move under direct fire												
21: Move over, around obstacles												
22: Camouflage individual equipment	86	52%										
23: Put on, wear, and remove M17 mask	75	72%										
24: Put on & were protective clothing	70	70%										
25: Perform CPR using one-man method	88	73%										
26: Put on field or pressure dressing	88	75%										
27: Navigate from point to point	81	65%	52	58%								
28: Establish observation post												
29: Conduct day & night surveillance												
30: Administer nerve agent antidote	75	58%										

Table 14b: Tasks Matched to Job Component with 50% or Greater Agreement

63B: Light Wheel Vehicle Mechanic
N=35

Project A Tasks	Task Category High	%	2nd	%	3rd	%	Job Activity High	%	2nd	%	3rd	%
1: Adjust cluch pedal	5	80%	51	63%								
2: Maintain assigned toolkit												
3: Repair electrical wiring	11	71%	51	51%			25	86%	21	51%		
4: Replace air hydraulic cylinder	5	83%	51	60%			26	63%	21	60%	34	57%
5: Replace radiator	5	83%	51	60%			21	63%	34	51%		
6: Replace service brakes	5	89%	51	63%			21	60%	26	60%	34	51%
7: Replace fuel pump	5	83%	51	63%			26	71%	21	63%	34	54%
8: Replace wheel bearings	5	80%	51	63%			21	69%	26	51%	34	51%
9: Tow disabled vehicle w/ wrecker	19	66%										
10: Perform annual PMCS	1	74%	51	51%			21	74%	11	63%		
11: Troubleshoot break system	51	66%	6	60%			26	77%	21	66%		
12: Troubleshoot electrical system	13	66%	51	54%			25	86%	21	69%		
13: Troubleshoot fuel system	6	74%	51	66%			26	83%	21	74%		
14: Troubleshoot service brake	6	74%	51	66%	5	51%	26	83%	21	77%		
15: Replace starter	5	69%	51	60%			34	69%	21	63%		
16: Administer first aid (Buddy-aid)	88	77%	75	54%			31	51%				
17: Camouflage equipment	86	80%										
18: Perform maintenance on M16A1	2	63%	51	54%			21	57%				
19: Determine grid coordinates	52	57%										
20: Determine magnetic azimuth	81	63%					21	51%				
21: load, reduce, clear M16A1	2	51%										
22: Use challenge and password							15	54%				
23: Put on M17 Mask w/ hood	75	69%										
24: Put on protective Clothing (MOPP)	75	71%										
25: Put on field/pressure dressing	88	77%										
26: Performe expedient repairs							26	57%				
27: Troubleshoot cooling system	51	60%	6	51%			26	77%	21	54%		
28: Troubleshoot engines	6	60%	51	60%			26	80%	21	54%		
29: Troubleshoot steering system	51	60%	6	54%			26	83%	21	57%		
30: Slave start disabled vehicle	51	54%					21	54%				

Table 14c: Tasks Matched to Job Component with 50% or Greater Agreement

71L: Administrative Specialist
N=39

Project A Tasks	Task Category High	%	2nd	%	3rd	%	Job Activity High	%	2nd	%	3rd	%
1: Type a basic comment to a DF	46	97%	47	56%			37	92%	21	56%		
2: Type a military letter	46	97%					37	92%	21	64%		
3: Type a joint messageform	46	97%	47	62%			37	90%	21	62%		
4: Prepare a requisition	47	85%					21	56%				
5: type military orders	46	97%					37	87%	21	67%		
6: Type a memorandum	46	97%					37	87%	21	64%		
7: Dispatch outgoing distribution	48	97%					21	59%				
8: Establish functional files	48	85%	51	54%			21	79%				
9: File documents/correspondence	48	85%					21	72%				
10: Type a second comment to a DF	46	95%					37	85%	21	62%		
11: Type straight copy material	46	97%					37	87%	21	51%		
12: Assemble correspondence	48	54%					21	62%				
13: Receipt classified material	48	56%					21	69%				
14: Safeguard FOUO material	48	54%					21	77%				
15: Perform maintenance on M16A1	2	79%										
16: Load, reduce, clear M16A1	2	62%										
17: Camouflage yourself & Equipment	82	69%	86	54%								
18: Practice Noice, light, discipline	82	74%										
19: Determine grid coordinates (GFS)	81	59%	52	54%								
20: Determine magnetic azimuth	81	59%										
21: Put on field/pressure dressing	88	95%					21	51%				
22: Maintain M17 mask w/ hood	75	79%					21	54%				
23: Put on M17 mask w/ hood	75	95%										
24: Put on Protective clothing (MOPP)	75	87%										
25: Know rights as POW	80	82%										
26: Administer nerve agent (self-aid)	75	67%										
27: Receive...office equipment	49	85%					21	51%				
28: Control expendable/non- supplies	49	82%										

Table 15a: Summary Table for Task Category Matching Data

	11B	63B	71L	11B/63B	11B/71L	63B/71L	11B/63B/71L
Number of Important Categories (Mean Greater Than or Equal to 3.00)	25	16	13	9	6	5	5
Number of Important Categories with 1 or More Tasks Matched (Agreement of at least 50%)	13	11	8	4	2	3	2
Percent of Total Important Categories With 1 or More Match	52%	69%	62%	44%	33%	60%	40%
Number of Important Categories with 2 or More Tasks Matched (Agreement of at least 50%)	4	6	6	3	1	1	1
Percent of Total Important Categories With 2 or More Tasks Matched	16%	38%	46%	33%	17%	20%	20%

Table 15b: Summary Table for Job Activity Matching Data

	11B	63B	71L	11B/63B	11B/71L	63B/71L	11B/63B/71L
Number of Important Activities (Mean Greater Than or Equal to 3.00)	14	19	10	7	5	4	4
Number of Important Activities with 1 or More Tasks Matched (Agreement of at least 50%)	2	4	3	--	--	1	--
Percent of Total Important Activities With 1 or More Match	14%	21%	33%	--	--	25%	--
Number of Important Activities with 2 or More Tasks Matched (Agreement of at least 50%)	--	4	3	--	--	1	--
Percent of Total Important Activities With 2 or More Tasks Matched	--	21%	33%	--	--	25%	--

Table 16 Task Category Questionnaire Evaluation

Task Category Definitions

1. How clear were the task category definitions?
(1=not at all clear, 7=very clear)

	N	Mean	S.D.
Overall	130	6.03	0.92
11B	56	6.00	0.87
63B	35	6.11	0.87
71L	39	6.00	1.05
NCOs	75	5.99	1.03
Officers	45	6.11	0.65
Civilians	10	6.00	1.15
FORSCOM	63	6.03	0.88
DOTD	67	6.03	0.97

2. How well did you understand the terms used in the task category definitions?
(1=there were many terms I did not understand, 7=there were no terms I did not understand)

	N	Mean	S.D.
Overall	130	6.42	0.74
11B	56	6.43	0.68
63B	35	6.34	0.94
71L	39	6.46	0.64
NCOs	75	6.44	0.76
Officers	45	6.42	0.58
Civilians	10	6.20	1.23
FORSCOM	63	6.51	0.62
DOTD	67	6.33	0.84

3. What percentage of your job was covered by these task categories?

	N	Mean	S.D.
Overall	127	81.10	26.55
11B	55	76.73	28.42
63B	34	91.18	15.72
71L	38	78.42	29.55
NCOs	74	74.86	30.03
Officers	44	88.86	18.70
Civilians	9	94.44	10.14
FORSCOM	62	78.55	29.30
DOTD	65	83.54	23.61

Table 16 continued

Task Category Ratings

1. How clear were the instructions for making the ratings? (1=not at all clear, 7=very clear)

	N	Mean	S.D.
Overall	123	6.24	1.09
11B	54	6.22	1.18
63B	33	6.55	0.67
71L	36	5.97	1.21
NCOs	72	6.22	1.13
Officers	41	6.37	0.70
Civilians	10	5.80	1.87
FORSCOM	58	6.34	0.89
DOTD	65	6.14	1.24

2. Were the ratings easy or difficult to make? (1=very difficult, 7=very easy)

	N	Mean	S.D.
Overall	123	3.92	1.77
11B	54	3.52	1.92
63B	33	4.24	1.46
71L	36	4.22	1.73
NCOs	72	3.94	1.86
Officers	41	4.07	1.47
Civilians	10	3.10	2.18
FORSCOM	58	3.83	1.79
DOTD	65	4.00	1.77

3. How confident did you feel about the accuracy of your ratings? (1=not at all confident, 7=very confident)

	N	Mean	S.D.
Overall	123	5.74	1.05
11B	54	5.72	1.11
63B	33	5.94	0.79
71L	36	5.58	1.16
NCOs	72	5.79	1.10
Officers	41	5.54	0.95
Civilians	10	6.20	0.92
FORSCOM	58	5.79	1.07
DOTD	65	5.69	1.03

Table 16 continued

General Evaluation

Which of the following would you change?

Overall	129	26%	Task Category Definitions
	129	15%	Task Category Rating Instructions and Procedures

Table 17 Job Activity Questionnaire Evaluation

Job Activity Definitions

1. How clear were the job activity definitions?
(1=not at all clear, 7=very clear)

	N	Mean	S.D.
Overall	117	5.40	1.25
11B	43	5.63	1.18
63B	35	5.23	1.19
71L	39	5.31	1.38
NCOs	70	5.16	1.42
Officers	42	5.79	0.84
Civilians	5	5.60	0.89
FORSCOM	65	5.58	1.21
DOTD	52	5.17	1.28

2. How well did you understand the terms used in the job activity definitions?
(1=there were many terms I did not understand, 7=there were no terms I did not understand)

	N	Mean	S.D.
Overall	117	6.03	1.06
11B	43	5.86	1.10
63B	35	6.00	1.21
71L	39	6.23	0.84
NCOs	70	5.97	1.19
Officers	42	6.10	0.85
Civilians	5	6.20	0.84
FORSCOM	65	6.18	0.95
DOTD	52	5.83	1.17

3. What percentage of your job was covered by these task categories?

	N	Mean	S.D.
Overall	114	77.54	24.87
11B	42	79.76	22.68
63B	35	77.43	25.36
71L	37	75.14	27.14
NCOs	68	71.76	26.93
Officers	41	85.61	18.98
Civilians	5	90.00	17.32
FORSCOM	64	77.97	24.64
DOTD	50	77.00	25.41

Table 17 continued

Job Activity Ratings

1. How clear were the instructions for making the ratings? (1=not at all clear, 7=very clear)

	N	Mean	S.D.
Overall	114	6.05	1.02
11B	40	5.97	0.95
63B	35	6.29	0.79
71L	39	5.92	1.24
NCOs	68	5.94	1.12
Officers	41	6.24	0.86
Civilians	5	6.00	0.71
FORSCOM	62	6.18	0.86
DOTD	52	5.90	1.18

2. Were the ratings easy or difficult to make? (1=very difficult, 7=very easy)

	N	Mean	S.D.
Overall	113	5.12	1.44
11B	40	5.17	1.53
63B	34	4.82	1.40
71L	39	5.31	1.36
NCOs	67	5.03	1.49
Officers	41	5.34	1.28
Civilians	5	4.40	1.95
FORSCOM	62	5.42	1.31
DOTD	51	4.75	1.51

3. How confident did you feel about the accuracy of your ratings? (1=not at all confident, 7=very confident)

	N	Mean	S.D.
Overall	114	5.50	1.00
11B	40	5.52	0.96
63B	35	5.29	1.05
71L	39	5.67	0.98
NCOs	68	5.49	1.06
Officers	41	5.51	0.90
Civilians	5	5.60	1.14
FORSCOM	62	5.61	1.03
DOTD	52	5.37	0.95

Table 17 continued

General Evaluation

Which of the following would you change?

Overall	117	36%	Job Activity Definitions
	117	21%	Job Activity Rating Instructions and Procedures

Table 18

Breakdown of Subjects by Method,

Rater Type, MOS, and Command

Variable	Cases
Method	
Importance	53
Validity	74
Rater Type	
Civilian	11
NCO	75
Officer	41
MOS	
11B	58
63B	32
71L	37
Command	
DOTD	62
FORSCOM	65
Total	127

Table 19

Rating of Performance Definition Detail by

Method, Rater Type, MOS, and Command

PERFDEF1: How would you describe the amount of detail and information provided in the performance definition?

1	2	3	4	5	6	7
Far too little			The right amount			Far too much

	Mean	Std Dev	Cases
Summary for Total Sample	4.55	.96	127
Summary by Method			
Importance	4.58	.95	53
Validity	4.53	.98	74
Summary by Rater Type			
Civilian	4.73	.47	11
NCO	4.55	1.09	75
Officer	4.51	.81	41
Summary by MOS			
11B	4.52	.98	58
63B	4.66	1.00	32
71L	4.51	.93	37
Summary by Command			
DOTD	4.42	.97	62
FORSCOM	4.68	.95	65

Table 20

Rating of Ranking Instruction Clarity by

Method, Rater Type, MOS, and Command

RANK1: How clear were the instructions for the ranking?

1	2	3	4	5	6	7
Not at all clear			Neither clear nor unclear			Very clear

	Mean	Std Dev	Cases
Summary for Total Sample	5.61	1.32	127
Summary by Method			
Importance	6.17	.78	53
Validity	5.22	1.47	74
Summary by Rater Type			
Civilian	4.82	1.47	11
NCO	5.53	1.43	75
Officer	5.98	.91	41
Summary by MOS			
11B	5.33	1.50	58
63B	5.84	1.19	32
71L	5.86	1.00	37
Summary by Command			
DOTD	5.82	1.29	62
FORSCOM	5.42	1.32	65

Table 21

Rating of Ranking Task Difficulty by

Method, Rater Type, MOS, and Command

RANK2: Was it easy or difficult to rank order the attributes?

1	2	3	4	5	6	7
Very difficult			Neither easy nor difficult			Very easy

	Mean	Std Dev	Cases
Summary for Total Sample	4.27	1.59	127
Summary by Method			
Importance	4.51	1.67	53
Validity	4.09	1.52	74
Summary by Rater Type			
Civilian	3.82	1.33	11
NCO	4.44	1.69	75
Officer	4.07	1.46	41
Summary by MOS			
11B	4.31	1.58	58
63B	3.97	1.67	32
71L	4.46	1.54	37
Summary by Command			
DOTD	4.31	1.62	62
FORSCOM	4.23	1.57	65

Table 22

Rating of Confidence in Ranking Accuracy by

Method, Rater Type, MOS, and Command

RANK3: How confident do you feel about the accuracy of your rankings?

1	2	3	4	5	6	7
Not at all confident			Somewhat confident			Very confident

	Mean	Std Dev	Cases
Summary for Total Sample	4.92	1.31	127
Summary by Method			
Importance	5.24	1.04	53
Validity	4.69	1.44	74
Summary by Rater Type			
Civilian	4.91	.94	11
NCO	4.89	1.42	75
Officer	4.98	1.21	41
Summary by MOS			
11B	4.86	1.28	58
63B	4.69	1.45	32
71L	5.22	1.23	37
Summary by Command			
DOTD	4.97	1.33	62
FORSCOM	4.88	1.30	65

Table 23

Correlations of Four Evaluation Ratings

for Total Sample and Method Subsamples

Correlations for Total Sample

	PERFDEF1	RANK1	RANK2	RANK3
PERFDEF1	1.0000			
RANK1	.0500	1.0000		
RANK2	.1770	.3720**	1.0000	
RANK3	.0408	.5152**	.6523**	1.0000

N = 127
* p < .001 , one-tailed

Correlations for Importance Subsample

	PERFDEF1	RANK1	RANK2	RANK3
PERFDEF1	1.0000			
RANK1	.0712	1.0000		
RANK2	.2813	.2872	1.0000	
RANK3	.0664	.5200**	.6928**	1.0000

N = 53
* p < .001 , one-tailed

Correlations for Validity Subsample

	PERFDEF1	RANK1	RANK2	RANK3
PERFDEF1	1.0000			
RANK1	.0338	1.0000		
RANK2	.0947	.4070**	1.0000	
RANK3	.0205	.4768**	.6391**	1.0000

N = 74
* p < .001 , one-tailed

Table 24

ATTDIFF: Think about the list of thirty-one attributes (below). Place a checkmark in front of any attributes that were difficult to judge or rank. Please check at least two attributes.

Attribute	Freq	Attribute	Freq
Verbal Ability	5	Work Orientation	7
Reasoning	8	Sociability	19
Number Ability	10	Cooperation/Stability	7
Spatial Ability	35	Energy	3
Closure	31	Conscientiousness	9
Mental Information Processing	8	Dominance/Confidence	17
Perceptual Speed and Accuracy	12	Interest in Using Tools and Machines	8
Memory	3	Interest in Rugged Activities	10
Mechanical Comprehension	8	Interest in Protective Services	24
Eye-Limb Coordination	13	Interest in Technical Activities	11
Precision	8	Interest in Science	23
Movement Judgment	13	Interest in Leadership	6
Hand and Finger Dexterity	10	Interest in Artistic Activities	25
Physical Strength	0	Interest in Efficiency and Organization	13
Physical Endurance	2		
Balance and Flexibility	12		
Involvement in Athletics	16		

Table 25

PERFDEF2: If you think the performance definition has too little detail and information, what would you add?

	Freq	%
No comment	116	91
Comment	11	9
Total	127	100

Comment	Freq
Right amount of detail	2
Make examples specific (to MOS)	2
Best part of today's survey	1
Add TAMMS portions	1
(Add) Offensive/defensive details	1
(Add) Easy to understand terminology	1
Make explanation clearer	1
Too vague--use scenarios	1
Closure is not clearly applicable[a]	1
Total	11

[a]Comment refers to attribute definitions, not performance definition.

Table 26

PERFDEF3: If you think the performance definition has too much detail and information, what would you delete?

	Freq	%
No comment	112	88
Comment	15	12
Total	127	100

Comment	Freq
Shorten, use fewer words	6
All important	1
(Delete) Examples	1
Rewrite Core Technical Proficiency (with regard to willingness)	1
Write specific definitions and examples of high and low performance	1
Too much can be read into this	1
Merge some attributes[a]	1
Too many attributes[a]	1
(Delete) Spatial Ability[a]	1
Combine Spatial Ability and Closure[a]	1
(Delete) Interest in Technical Activities[a]	1
(Delete) Interest in Artistic Activities[a]	1
(Delete) Descriptions of high, average, and low levels on attributes[a]	1
Movement Judgment has limited importance[a]	1
Total	19[b]

[a]Comment refers to attribute definitions, not performance definition.
[b]Some subjects provided more than one comment.

Table 27

ATTADD: Are there any attributes you would add?

	Freq	%
No comment	109	86
Comment	18	14
Total	127	100

Comment	Freq
No additions necessary	4
Temperament/Attitude	
Integrity/Judgment/Moral Discretion	2
Coping with Stress	1
Replace Work Orientation with Dedication to Duty	1
Military Bearing	1
More temperament attributes	1
Loyalty to Country	1
Optimism	1
Value of Human Life	1
Sense of Humor	1
Interests	
Interest in Values	1
Interest in Family	1
Interest in Writing	1
Interest in Business	1
Interest in History	1
Cognitive Abilities	
Attention to Detail	1
Ability to Communicate (e.g., bilingual)	1
Reading Comprehension	1
Research Ability	1
Coordination & Physical Abilities	
Team Sports Player	1
Team Sports Captain	1
Miscellaneous	
Religion	1
Ability to Kill (mentally)	1
Total	27[a]

[a]Some subjects provided more than one comment.

Table 28

ATTDEL: Are there any attributes you would delete?

	Freq	%
No comment	99	78
Comment	28	22
Total	127	100

Comment	Freq
Yes, some attributes should be deleted	4
Interests	
Interest in Artistic Activities	16
Interest in Science	6
Interest in Protective Services	2
Interest in Rugged Activities	2
Interest in Technical Activities	1
Coitive Abilities	
Closure	3
Spatial Ability	1
Combine Reasoning, Spatial Ability, Closure, and Mental Information Processing into one attribute	1
Temperament/Attitude	
Cooperation (need only Stability)	1
Dominance (need only Confidence)	1
Coordination & Physical Abilities	
Involvement in Athletics	2
Total	40[a]

[a]Some subjects provided more than one comment.

Table 29

General Evaluation Summary

GENERAL EVALUATION: Think about the two tasks you completed, attribute judgments and ranking.

GENEVAL1: Which was easier?

	Freq	%
Judgment	63	50
Ranking	51	40
Both[a]	2	2
Neither[a]	11	9
Total	127	101[b]

GENEVAL2: Which provides more information about the MOS?

	Freq	%
Judgment	80	63
Ranking	35	28
Both[a]	2	2
Neither[a]	10	8
Total	127	101[b]

GENEVAL3: If the Army was going to use one of these methods to guide selection and placement of soldiers in MOS, which would you prefer?

	Freq	%
Judgment	72	57
Ranking	42	33
Both[a]	3	2
Neither[a]	10	8
Total	127	100

[a]This response option was not provided on the form; some subjects added it.

[b]Percentages do not sum to 100 due to rounding error.

Table 30. Cross-Tabulation of MOS and Rater Type for Importance Rating Method

	Rater Type			
	NCO	Officer	Civilian	Total
MOS				
11B	21	2	1	24
63B	4	7	3	14
71L	17	16	1	34
Total	42	25	5	72

Table 31. Cross-Tablulation of MOS and Rater Type for Validity Rating Method

	Rater Type			
	NCO	Officer	Civilian	Total
MOS				
11B	16	16	6	38
63B	18	6	0	24
71L	16	12	1	29
Total	50	34	7	91

Table 32. **ANOVA of Validity Ratings for Core Technical Performance Area**

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	148.58	5.34	<.01
Type	1	13.74	0.49	0.49
MOS x Type	2	14.47	0.52	0.60
Error	65	27.84		
Within Subjects:				
Attributes	30	54.51	23.31	<.001
Attributes x MOS	60	19.85	8.49	<.001
Attributes x Type	30	3.50	1.50	0.04
Attributes x MOS x Type	60	5.78	2.47	<.001
Error	1950	2.34		

Note: Type includes NCOs and Officers only.

Table 33. ANOVA of Validity Ratings for General Soldiering Performance Area

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	65.29	2.04	0.14
Type	1	34.67	1.08	0.30
MOS x Type	2	25.56	.80	0.46
Error	65	32.07		
Within Subjects:				
Attributes	30	42.51	20.97	<.001
Attributes x MOS	60	2.86	1.41	.02
Attributes x Type	30	4.04	1.99	<.001
Attributes x MOS x Type	60	3.32	1.64	<.01
Error	1950	2.03		

Note: Type includes NCOs and Officers only.

Table 34. **ANOVA of Validity Ratings for Effort/Leadership Performance Area**

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	262.75	9.72	<.001
Type	1	1.70	.06	0.80
MOS x Type	2	38.88	1.44	0.25
Error	65	27.04		
Within Subjects:				
Attributes	30	79.88	34.01	<.001
Attributes x MOS	60	2.45	1.04	0.39
Attributes x Type	30	8.26	3.52	<.001
Attributes x MOS x Type	60	4.82	2.05	<.001
Error	1950	2.35		

Note: Type includes NCOs and Officers only.

Table 35. **ANOVA of Validity Ratings for Personal Discipline Performance Area**

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	286.18	6.82	<.01
Type	1	177.75	4.24	0.04
MOS x Type	2	297.76	7.10	<.01
Error	64	41.96		
Within Subjects:				
Attributes	30	70.13	27.03	<.001
Attributes x MOS	60	3.25	1.25	0.09
Attributes x Type	30	8.43	3.25	<.001
Attributes x MOS x Type	60	3.13	1.21	0.14
Error	1920	2.59		

Note: Type includes NCOs and Officers only.

Table 36. ANOVA of Validity Ratings for Physical Fitness/Military Bearing Performance Area

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	188.67	4.70	<.01
Type	1	230.43	5.74	0.02
MOS x Type	2	251.47	6.27	<.01
Error	65	40.13		
Within Subjects:				
Attributes	30	136.99	46.24	<.001
Attributes x MOS	60	4.38	1.48	<.01
Attributes x Type	30	8.65	2.92	<.001
Attributes x MOS x Type	60	4.87	1.64	<.01
Error	1950	2.96		

Note: Type includes NCOs and Officers only.

Table 37. ANOVA of Validity Rankings

Source	*df*	*MS*	*F*	*p*
Between Subjects:				
MOS	2	1.18	1.12	0.33
Type	1	0.86	0.82	0.37
MOS x Type	2	1.16	1.10	0.34
Error	61	1.05		
Within Subjects:				
Attributes	30	1331.13	29.41	<.001
Attributes x MOS	60	382.76	8.46	<.001
Attributes x Type	30	70.76	1.56	0.03
Attributes x MOS x Type	60	57.33	1.27	0.08
Error	1830	45.26		

Note: Type includes NCOs and Officers only.

Table 38. Reliabilities of Validity Ratings by MOS, Rater Type, and Criterion

			Performance Area					
		N[a]	Core	General	Eff	Disc	Fit	RNK[b]
11B	NCOs	15	.76	.75	.79	.73	.89	.92 (14)
	Officers	13	.78	.83	.94	.93	.95	.93 (15)
	NCOs & Officers	28	.76	.78	.88	.85	.92	.93 (29)
63B	NCOs	17	.79	.73	.81	.77	.84	.77 (13)
	Officers	6	N/A	N/A	N/A	N/A	N/A	N/A (3)
	NCOs & Officers	21	.85	.79	.84	.81	.88	.83 (16)
71L	NCOs	12	.93	.86	.84	.81	.89	.96 (13)
	Officers	10	.94	.86	.91	.91	.94	.95 (10)
	NCOs & Officers	22	.93	.84	.87	.87	.91	.95 (23)

Note: Reliabilities adjusted to $K = 15$.

[a] These sample sizes refer to the performance area analyses.

[b] Sample sizes for the ranking analyses are in parentheses.

Table 39. Reliabilities of Importance Ratings by MOS, Rater Type, and Criterion

			Performance Area					
		N[a]	Core	General	Eff	Disc	Fit	RNK[b]
11B	NCOs	17	.89	.89	.86	.83	.92	.94 (20)
	Officers	2	N/A	N/A	N/A	N/A	N/A	N/A (1)
	NCOs & Officers	18	.90	.90	.87	.84	.93	.94 (21)
63B	NCOs	4	N/A	N/A	N/A	N/A	N/A	N/A (3)
	Officers	7	N/A	N/A	N/A	N/A	N/A	N/A (5)
	NCOs & Officers	8	.91	.88	.91	.86	.91	.94 (8)
71L	NCOs	16	.95	.87	.91	.85	.91	.96 (13)
	Officers	14	.95	.89	.92	.88	.95	.96 (13)
	NCOs & Officers	30	.95	.87	.91	.87	.93	.96 (26)

Note: Reliabilities adjusted to $K = 15$.

[a] These sample sizes refer to the performance area analyses.

[b] Sample sizes for the ranking analyses are in parentheses.

Table 40. Correlations of Mean Importance and Mean Validity Ratings by MOS for Core Technical Proficiency, Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.41	1.00				
71L Imp	.38	.34	1.00			
11B Val	.81	.51	.45	1.00		
63B Val	.45	.88	.35	.52	1.00	
71L Val	.39	.37	.98	.49	.35	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 41. Correlations of Mean Importance and Mean Validity Ratings by MOS for General Soldier Proficiency, Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.71	1.00				
71L Imp	.81	.87	1.00			
11B Val	.80	.84	.90	1.00		
63B Val	.75	.83	.85	.87	1.00	
71L Val	.82	.82	.89	.87	.78	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 42. Correlations of Mean Importance and Mean Validity Ratings by MOS for Effort and Leadership, Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.71	1.00				
71L Imp	.72	.92	1.00			
11B Val	.80	.87	.93	1.00		
63B Val	.77	.90	.89	.93	1.00	
71L Val	.81	.86	.92	.92	.86	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 43. Correlations of Mean Importance and Mean Validity Ratings by MOS for Personal Discipline, Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.81	1.00				
71L Imp	.77	.85	1.00			
11B Val	.85	.86	.87	1.00		
63B Val	.81	.89	.89	.92	1.00	
71L Val	.83	.86	.93	.89	.90	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 44. Correlations of Mean Importance and Mean Validity Ratings by MOS for Physical Fitness/Military Bearing, Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.89	1.00				
71L Imp	.95	.94	1.00			
11B Val	.94	.89	.96	1.00		
63B Val	.91	.92	.94	.94	1.00	
71L Val	.93	.91	.95	.91	.91	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 45. Correlations of Mean Importance and Mean Validity Ranks Computed Across the 31 Attributes

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.20	1.00				
71L Imp	.60	.27	1.00			
11B Val	.95	.31	.56	1.00		
63B Val	.48	.84	.53	.54	1.00	
71L Val	.60	.29	.98	.55	.54	1.00

Note: Mean ranks are based on the combined NCO and Officer groups.

Table 46. Correlations of Mean Importance and Mean Validity Ratings, Computed Across the 155 Attribute x Performance Area Combinations

	11B Imp	63B Imp	71L Imp	11B Val	63B Val	71L Val
11B Imp	1.00					
63B Imp	.75	1.00				
71L Imp	.75	.77	1.00			
11B Val	.88	.81	.82	1.00		
63B Val	.78	.89	.79	.87	1.00	
71L Val	.77	.76	.95	.81	.76	1.00

Note: Mean ratings are based on the combined NCO and Officer groups.

Table 47. Correlations of Mean Importance and Mean Validity Rankings with Mean Importance and Mean Validity Ratings in Each of the Five Performance Areas for 11B: Means are based on NCO and Officer groups.

	Mean Importance Ranks	Mean Validity Ranks
Mean Importance:		
Core Technical	.864	.827
General Soldier	.868	.843
Effort/Leadership	.823	.787
Personal Discipline	.790	.808
Fitness/Bearing	.574	.613
Mean Validity:		
Core Technical	.755	.789
General Soldier	.754	.815
Effort/Leadership	.627	.685
Personal Discipline	.473	.564
Fitness/Bearing	.568	.630

Note: N = 31 for each of the correlations of means (number of attributes). Correlations have been appropriately reflected.

Table 48. Correlations of Mean Importance and Mean Validity Rankings with Mean Importance and Mean Validity Ratings in Each of the Five Performance Areas for 63B: Means are based on NCO and Officer groups.

	Mean Importance Ranks	Mean Validity Ranks
Mean Importance:		
Core Technical	.919	.835
General Soldier	.344	.473
Effort/Leadership	.086	.201
Personal Discipline	.166	.294
Fitness/Bearing	.178	.225
Mean Validity:		
Core Technical	.920	.813
General Soldier	.310	.329
Effort/Leadership	.070	.152
Personal Discipline	.013	.056
Fitness/Bearing	-.004	.006

Note: N = 31 for each of the correlations of means (number of attributes). Correlations have been appropriately reflected.

Table 49. Correlations of Mean Importance and Mean Validity Rankings with Mean Importance and Mean Validity Ratings in Each of the Five Performance Areas for 71L: Means are based on NCO and Officer groups.

	Mean Importance Ranks	Mean Validity Ranks
Mean Importance:		
Core Technical	.952	.955
General Soldier	.657	.633
Effort/Leadership	.670	.606
Personal Discipline	.604	.534
Fitness/Bearing	-.027	-.028
Mean Validity:		
Core Technical	.944	.952
General Soldier	.572	.569
Effort/Leadership	.659	.591
Personal Discipline	.528	.461
Fitness/Bearing	-.097	-.083

Note: N = 31 for each of the correlations of means (number of attributes). Correlations have been appropriately reflected.

Table 50. Mean Importance Ratings of Attributes for Core Technical Proficiency: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	6.19	0.81	21	4.63	2.39	8	6.23	1.41	30
Reason	6.38	1.02	21	5.12	1.64	8	5.57	1.59	30
Number	4.53	2.06	19	3.12	1.64	8	3.70	1.66	30
* Spatial	5.76	1.18	21	5.38	2.26	8	3.23	2.14	30
* Closure	5.48	1.12	21	2.87	1.81	8	5.80	2.01	30
Info Process	6.81	1.89	21	4.88	2.17	8	6.03	1.54	30
PSA	6.14	1.39	21	4.88	2.30	8	6.00	1.84	30
Memory	7.10	0.89	21	5.63	2.00	8	6.33	1.24	30
* Mech Comp	4.62	1.77	21	7.13	1.25	8	3.13	2.37	30
Eye-Limb Coord	6.67	1.24	21	5.75	2.19	8	4.77	2.33	30
* Precision	7.19	0.81	21	5.88	1.46	8	4.07	2.53	30
* Movement Judge	6.14	1.39	21	4.50	2.93	8	1.83	1.82	30
Dexterity	6.52	1.29	21	7.13	1.13	8	6.20	2.16	30
* Strength	7.00	1.10	21	5.37	1.60	8	2.57	1.79	30
* Endurance	7.19	1.03	21	5.25	1.83	8	2.73	1.76	30
* Balance	6.10	1.00	21	5.38	1.85	8	2.17	1.97	30
* Athletics	5.62	1.50	21	1.87	2.23	8	1.53	1.85	30
Work Orient	6.14	1.31	21	5.38	2.00	8	5.93	1.76	30
Sociability	5.24	1.79	21	3.25	2.43	8	4.67	1.69	30
Cooperation	4.90	2.07	21	3.75	2.05	8	5.83	1.68	30
Energy	5.90	1.34	21	4.13	2.17	8	5.33	1.75	30
Conscientious	5.90	1.37	21	4.75	1.91	8	6.20	1.65	30
Dominance	6.24	1.41	21	4.25	2.60	8	4.90	1.71	30
* Int. Tools	4.89	2.02	19	7.63	0.74	8	2.00	2.35	30
* Int. Rugged	6.20	1.40	20	2.88	2.03	8	0.87	1.25	30
* Int. Protect	5.05	2.44	20	1.63	2.07	8	1.10	1.49	30
Int. Tech	4.10	1.77	20	5.75	2.38	8	3.40	2.80	30
Int. Science	2.50	2.24	20	3.50	2.27	8	1.47	2.00	30
Int. Leader	5.50	2.06	20	3.63	2.92	8	3.80	2.38	30
Int. Art	2.30	2.49	20	0.63	1.06	8	1.23	1.94	30
Int. Organize	4.45	2.56	20	5.38	1.51	8	5.60	2.08	30

Table 51. Mean Importance Ratings of Attributes for General Soldiering Proficiency: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	5.95	1.07	21	3.75	1.83	8	4.57	1.63	30
Reason	6.00	1.26	21	4.38	2.07	8	4.93	1.51	30
Number	4.21	1.93	19	2.88	1.73	8	3.47	1.57	30
* Spatial	5.57	1.08	21	2.50	1.07	8	4.30	1.78	30
Closure	5.24	1.09	21	2.88	2.10	8	4.03	1.65	30
Info Process	6.67	1.24	21	5.13	1.36	8	5.53	1.74	30
PSA	5.95	1.28	21	4.38	1.77	8	5.13	1.48	30
Memory	6.90	1.14	21	5.75	1.83	8	5.80	1.19	30
Mech Comp	4.76	1.48	21	3.00	2.39	8	4.20	2.17	30
Eye-Limb Coord	6.29	1.23	21	4.88	1.96	8	4.70	1.82	30
Precision	6.62	1.12	21	4.13	1.46	8	4.70	2.28	30
Movement Judge	5.90	1.30	21	4.00	1.77	8	3.87	2.24	30
Dexterity	6.33	1.20	21	4.13	1.73	8	4.60	1.85	30
Strength	6.71	1.15	21	5.13	1.55	8	4.97	1.52	30
Endurance	6.90	1.18	21	5.38	1.60	8	5.57	1.41	30
Balance	5.81	0.98	21	4.25	1.49	8	4.77	1.68	30
* Athletics	5.52	1.54	21	2.63	1.92	8	3.47	2.10	30
Work Orient	5.86	1.31	21	5.25	1.83	8	5.47	1.70	30
Sociability	5.00	1.73	21	4.25	1.75	8	3.93	1.84	30
Cooperation	4.86	2.06	21	4.13	1.89	8	4.97	1.81	30
Energy	5.57	1.36	21	4.25	1.49	8	5.63	1.43	30
Conscientious	5.71	1.15	21	5.13	1.96	8	5.97	1.38	30
Dominance	6.05	1.47	21	4.75	1.67	8	5.70	1.44	30
Int. Tools	4.68	1.95	19	3.38	2.72	8	3.47	1.87	30
Int. Rugged	5.85	1.76	20	4.38	1.06	8	4.43	1.87	30
Int. Protect	4.70	2.23	20	3.25	1.39	8	3.17	2.41	30
Int. Tech	4.00	1.65	20	3.88	2.36	8	3.43	2.36	30
Int. Science	2.55	2.09	20	1.63	1.92	8	2.00	1.97	30
Int. Leader	5.25	2.15	20	6.00	1.77	8	5.50	1.91	30
Int. Art	2.20	2.44	20	0.50	1.07	8	1.07	1.46	30
Int. Organize	4.30	2.54	20	5.50	1.77	8	4.60	2.01	30

Table 52. Mean Importance Ratings of Attributes for Effort and Leadership: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	6.19	1.97	21	5.00	2.20	8	5.33	1.86	30
Reason	6.10	1.45	21	4.88	2.23	8	5.57	1.55	30
* Number	4.95	1.31	19	1.88	1.64	8	3.00	1.62	30
Spatial	5.38	1.40	21	3.13	2.17	8	3.17	2.41	30
Closure	5.19	1.50	21	3.63	2.07	8	3.53	2.16	30
Info Process	6.24	1.51	21	4.88	1.64	8	4.93	1.86	30
PSA	5.52	1.54	21	3.25	1.91	8	4.70	1.78	30
Memory	6.62	1.20	21	4.63	1.85	8	5.40	1.45	30
Mech Comp	4.52	1.54	21	3.00	2.27	8	3.23	1.98	30
* Eye-Limb Coord	5.76	1.51	21	4.00	2.39	8	3.10	1.75	30
* Precision	6.29	1.35	21	3.50	1.69	8	3.07	2.32	30
* Movement Judge	5.62	1.56	21	2.75	2.05	8	2.50	1.89	30
Dexterity	4.24	2.45	21	2.75	2.82	8	2.37	1.85	30
* Strength	6.33	1.85	21	3.75	2.31	8	4.00	2.08	30
Endurance	6.62	1.72	21	5.38	0.74	8	4.77	2.10	30
* Balance	5.71	1.71	21	2.75	1.67	8	2.80	2.20	30
Athletics	5.71	1.59	21	3.87	1.81	8	3.23	2.13	30
Work Orient	6.19	1.60	21	5.62	1.69	8	5.93	1.68	30
Sociability	5.76	1.70	21	5.88	1.73	8	5.00	1.95	30
Cooperation	5.81	1.83	21	6.12	1.46	8	6.03	1.16	30
Energy	6.14	1.06	21	5.50	1.85	8	6.17	1.39	30
Conscientious	6.33	1.24	21	5.75	1.58	8	6.13	1.50	30
Dominance	6.67	1.28	21	6.50	1.69	8	6.53	1.20	30
Int. Tools	4.11	1.85	19	3.00	2.62	8	2.73	2.20	30
* Int. Rugged	5.80	1.44	20	2.38	2.20	8	3.07	2.12	30
* Int. Protect	6.05	1.73	20	3.13	1.55	8	3.50	2.47	30
Int. Tech	4.10	1.52	20	3.25	2.96	8	2.63	2.17	30
Int. Science	3.45	2.37	20	1.63	2.26	8	2.27	1.89	30
Int. Leader	6.80	1.24	20	6.88	1.36	8	6.27	1.93	30
Int. Art	2.60	2.44	20	1.00	1.60	8	1.57	1.70	30
Int. Organize	5.10	2.57	20	5.88	1.73	8	5.03	1.96	30

Table 53. Mean Importance Ratings of Attributes for Personal Discipline: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	4.86	1.35	21	2.88	2.64	8	2.90	2.40	30
Reason	4.90	1.41	21	4.13	2.23	8	3.83	1.98	30
* Number	3.58	1.92	19	1.00	1.41	8	2.37	2.13	30
Spatial	3.86	2.37	21	2.00	2.56	8	2.07	1.95	30
Closure	3.43	2.25	21	2.50	2.33	8	2.33	2.04	30
Info Process	5.38	1.69	21	3.38	1.85	8	3.53	2.08	30
PSA	4.57	2.31	21	3.25	2.31	8	3.10	1.86	30
Memory	5.76	1.61	21	3.88	1.64	8	4.30	1.88	30
Mech Comp	3.67	2.35	21	1.75	2.55	8	1.97	1.85	30
Eye-Limb Coord	4.62	2.91	21	3.25	2.76	8	2.30	2.14	30
* Precision	5.10	2.32	21	3.13	2.23	8	2.47	2.36	30
* Movement Judge	4.67	1.91	21	1.88	1.81	8	1.83	1.74	30
Dexterity	4.14	2.46	21	2.00	2.88	8	2.00	1.89	30
* Strength	6.10	2.17	21	3.63	2.45	8	2.70	2.10	30
* Endurance	6.48	2.04	21	4.13	1.89	8	3.63	2.17	30
* Balance	5.19	1.89	21	2.38	1.92	8	2.20	1.79	30
* Athletics	5.62	1.86	21	2.63	2.26	8	3.27	2.30	30
Work Orient	6.19	1.40	21	5.50	1.60	8	5.57	1.85	30
Sociability	4.86	1.74	21	2.50	1.93	8	3.30	2.28	30
Cooperation	5.48	1.99	21	4.50	2.33	8	4.47	2.16	30
Energy	5.52	1.47	21	3.63	1.85	8	5.20	1.92	30
Conscientious	6.00	1.70	21	5.00	2.20	8	5.57	2.30	30
Dominance	5.81	1.60	21	3.38	2.07	8	4.83	2.13	30
Int. Tools	3.42	2.06	19	2.00	2.67	8	1.53	1.59	30
* Int. Rugged	4.50	2.16	20	1.50	1.41	8	2.13	2.03	30
Int. Protect	5.05	1.90	20	3.25	2.19	8	2.63	2.36	30
Int. Tech	3.55	2.11	20	2.13	2.64	8	1.90	1.92	30
Int. Science	2.70	2.47	20	1.38	1.69	8	1.67	1.77	30
Int. Leader	5.95	2.06	20	5.00	2.62	8	4.17	2.41	30
Int. Art	2.60	2.46	20	0.75	1.16	8	1.50	1.80	30
Int. Organize	5.00	2.58	20	4.87	2.47	8	3.90	2.66	30

Table 54. Mean Importance Ratings of Attributes for Physical Fitness/Military Bearing: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	2.38	2.13	21	2.00	1.69	8	1.83	1.82	30
Reason	3.14	2.13	21	2.63	2.13	8	2.07	1.91	30
Number	2.84	2.29	19	1.00	1.41	8	1.03	1.30	30
Spatial	2.05	2.09	21	1.00	1.41	8	1.10	1.65	30
Closure	1.90	2.12	21	1.50	1.77	8	1.17	1.66	30
Info Process	4.14	2.43	21	2.25	2.25	8	1.87	1.87	30
PSA	3.29	2.65	21	2.00	2.62	8	1.60	1.90	30
Memory	3.81	2.36	21	2.88	1.64	8	1.67	1.52	30
Mech Comp	2.52	1.99	21	1.38	2.33	8	1.17	1.46	30
Eye-Limb Coord	5.81	2.02	21	4.63	2.33	8	4.23	2.50	30
Precision	5.29	1.82	21	3.88	1.64	8	3.13	2.47	30
Movement Judge	4.00	2.12	21	2.38	1.92	8	1.87	1.80	30
Dexterity	3.86	2.33	21	3.25	2.87	8	2.60	2.40	30
Strength	7.43	0.98	21	5.75	1.58	8	6.20	2.04	30
Endurance	7.57	0.75	21	6.38	1.41	8	6.77	1.65	30
Balance	6.43	1.43	21	4.63	1.69	8	4.47	2.43	30
Athletics	6.95	1.50	21	4.63	1.92	8	5.70	2.17	30
Work Orient	5.33	1.83	21	5.13	1.64	8	4.17	2.45	30
Sociability	2.90	2.51	21	2.00	1.93	8	1.70	1.97	30
Cooperation	3.62	2.42	21	1.75	1.75	8	2.10	2.14	30
Energy	5.55	1.39	20	4.13	1.25	8	4.77	2.57	30
Conscientious	4.29	2.41	21	3.63	2.62	8	3.63	2.59	30
* Dominance	6.19	1.54	21	3.00	1.77	8	3.97	2.31	30
Int. Tools	3.11	2.49	19	1.75	2.19	8	1.20	1.69	30
Int. Rugged	5.35	2.50	20	2.88	2.36	8	3.77	2.21	30
* Int. Protect	4.20	2.69	20	0.88	1.36	8	1.73	2.05	30
Int. Tech	2.20	2.04	20	1.25	2.05	8	1.00	1.39	30
Int. Science	1.85	2.35	20	0.75	1.39	8	0.63	0.96	30
* Int. Leader	5.05	2.52	20	2.50	2.27	8	2.67	2.62	30
Int. Art	1.70	2.34	20	0.50	1.07	8	0.50	0.68	30
Int. Organize	3.55	2.67	20	2.88	1.55	8	1.67	1.90	30

Table 55. Mean Importance Rankings of Attributes: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	7.00	6.31	21	17.00	5.48	8	3.50	2.66	26
Reason	8.33	8.11	21	12.38	6.21	8	5.73	3.73	26
Number	21.10	6.95	21	19.78	6.67	9	13.88	6.94	26
Spatial	18.76	7.35	21	9.25	6.80	8	16.19	6.68	26
Closure	22.52	4.70	21	22.44	6.64	9	11.38	7.33	26
Info Process	8.95	5.84	21	15.75	7.23	8	5.92	3.98	26
PSA	13.76	5.83	21	16.44	8.28	9	9.77	6.05	26
Memory	9.95	6.24	21	10.00	3.71	9	7.27	4.34	26
Mech Comp	21.95	4.89	21	2.13	1.89	8	22.50	5.81	26
Eye-Limb Coord	12.05	4.88	21	10.67	8.00	9	14.81	8.52	26
Precision	12.86	5.34	21	12.50	6.02	8	16.92	5.14	26
Movement Judge	15.33	4.88	21	16.56	7.33	9	22.04	3.39	26
Dexterity	15.76	4.97	21	8.67	3.87	9	9.81	6.36	26
Strength	6.14	5.01	21	13.00	6.99	8	19.46	5.29	26
Endurance	3.71	2.19	21	14.50	4.24	8	18.65	5.63	26
Balance	13.19	6.56	21	15.56	7.32	9	23.35	4.60	26
Athletics	19.05	9.13	21	25.75	5.50	8	24.50	4.55	26
Work Orient	13.10	6.58	21	8.88	7.77	8	6.77	4.39	26
Sociability	14.67	8.52	21	23.67	7.12	9	14.19	6.49	26
Cooperation	13.52	8.44	21	18.50	6.61	8	9.77	5.49	26
Energy	14.76	7.57	21	12.67	5.57	9	13.77	4.74	26
Conscientious	13.10	8.20	21	14.13	7.61	8	7.35	4.58	26
Dominance	13.29	6.41	21	17.56	7.37	9	15.54	5.47	26
Int. Tools	25.00	5.09	21	4.11	3.02	9	25.31	4.64	26
Int. Rugged	16.95	9.18	21	17.00	8.65	8	26.23	3.79	26
Int. Protect	23.10	6.30	21	24.89	7.32	9	25.54	4.43	26
Int. Tech	26.33	3.57	21	7.38	7.71	8	21.08	8.56	26
Int. Science	28.10	4.37	21	23.44	6.69	9	27.04	4.45	26
Int. Leader	12.90	7.44	21	24.00	3.93	8	17.38	5.76	26
Int. Art	30.14	2.10	21	30.38	0.74	8	28.00	4.71	26
Int. Organize	20.14	7.22	21	14.11	8.96	9	12.23	7.58	26

Table 56. Mean Validity Ratings of Attributes for Core Technical Proficiency: 11B, 63B, and 71L NCOs and Officers Combined

Variable	11B Mean	11B Std Dev	11B N	63B Mean	63B Std Dev	63B N	71L Mean	71L Std Dev	71L N
Verbal	5.27	1.20	30	4.52	1.66	21	6.33	0.96	24
Reason	5.47	1.61	30	5.86	1.31	21	5.56	1.45	25
Number	4.55	1.48	29	4.10	1.48	21	4.52	1.19	25
Spatial	4.97	1.73	30	5.48	1.57	21	3.76	1.96	25
Closure	4.46	1.88	28	4.62	1.80	21	5.40	1.63	25
Info Process	5.83	1.56	30	5.24	1.22	21	5.96	1.43	25
PSA	5.76	1.15	29	5.19	0.98	21	6.08	1.61	25
Memory	6.53	1.11	30	5.90	1.37	21	6.36	1.04	25
* Mech Comp	5.59	1.35	29	6.90	1.09	21	3.54	2.11	24
Eye-Limb Coord	5.40	1.57	30	6.19	1.17	21	5.04	2.03	25
Precision	5.73	1.36	30	5.14	1.53	21	4.92	2.27	25
* Movement Judge	5.79	1.35	29	4.29	1.27	21	2.96	2.09	25
Dexterity	4.90	1.79	30	5.95	1.77	21	6.16	1.43	25
* Strength	5.30	2.09	30	6.05	1.60	21	2.72	1.62	25
* Endurance	5.63	2.16	30	5.67	1.39	21	3.04	1.84	25
* Balance	4.73	1.89	30	5.62	1.50	21	2.16	1.72	25
* Athletics	4.76	2.37	29	3.29	2.24	21	1.56	2.02	25
Work Orient	5.33	2.14	30	6.33	1.53	21	6.12	1.24	25
Sociability	4.37	2.08	30	3.76	2.41	21	4.20	2.42	25
Cooperation	5.03	2.11	29	4.81	1.69	21	5.12	1.79	25
Energy	5.07	2.23	29	5.76	1.48	21	5.28	1.86	25
Conscientious	5.66	1.90	29	5.67	1.59	21	5.80	1.71	25
Dominance	5.21	1.70	29	5.00	2.00	21	4.56	1.78	25
* Int. Tools	4.76	2.08	29	7.19	0.93	21	2.16	2.12	25
* Int. Rugged	5.33	2.14	30	4.48	2.11	21	1.88	2.24	25
* Int. Protect	4.23	2.19	30	4.24	2.77	21	1.48	1.58	25
Int. Tech	4.27	2.26	30	5.71	1.35	21	3.75	2.59	24
Int. Science	3.59	2.64	29	3.43	2.64	21	2.00	1.79	24
Int. Leader	5.93	1.46	29	4.62	2.20	21	4.16	2.01	25
Int. Art	2.69	2.38	29	2.81	2.14	21	1.88	1.96	24
Int. Organize	4.83	1.98	29	5.57	2.11	21	6.13	1.39	24

Table 57. Mean Validity Ratings of Attributes for General Soldiering Proficiency: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	4.67	1.24	30	4.38	1.36	21	4.63	1.35	24
Reason	5.10	1.45	30	5.00	1.18	21	5.16	1.37	25
Number	3.86	1.62	29	4.00	1.34	21	3.68	1.18	25
Spatial	4.50	1.68	30	4.62	1.07	21	4.08	1.85	25
Closure	4.14	1.58	28	4.29	1.49	21	4.12	1.83	25
Info Process	5.33	1.52	30	5.48	1.60	21	5.48	1.26	25
PSA	5.34	1.17	29	5.24	1.55	21	4.80	1.58	25
Memory	6.30	1.02	30	5.67	1.93	21	5.92	1.44	25
Mech Comp	5.41	1.18	29	5.19	2.04	21	4.46	1.50	24
Eye-Limb Coord	4.97	1.33	30	5.95	1.40	21	4.92	1.89	25
Precision	5.53	1.41	30	5.62	1.40	21	5.24	1.88	25
Movement Judge	4.83	1.39	29	5.05	1.69	21	4.52	2.22	25
Dexterity	4.83	1.66	30	5.24	1.37	21	4.64	2.33	25
Strength	5.30	1.56	30	5.95	1.43	21	4.84	1.72	25
Endurance	5.60	1.71	30	6.29	1.31	21	5.32	1.82	25
Balance	4.73	1.62	30	4.76	1.70	21	4.68	1.73	25
Athletics	4.55	2.10	29	4.90	2.12	21	3.36	2.25	25
Work Orient	5.43	2.06	30	6.33	1.53	21	4.72	2.13	25
Sociability	4.57	1.85	30	4.90	2.39	21	2.64	2.23	25
Cooperation	5.17	1.73	29	5.52	1.44	21	4.40	2.10	25
Energy	5.48	1.64	29	6.00	1.90	21	4.92	2.04	25
Conscientious	5.45	1.70	29	6.05	1.72	21	4.92	2.02	25
Dominance	5.17	1.42	29	6.05	1.60	21	4.76	1.98	25
Int. Tools	4.41	1.90	29	4.14	2.24	21	2.80	1.80	25
Int. Rugged	5.23	1.81	30	5.76	1.76	21	4.44	2.38	25
Int. Protect	4.57	1.81	30	5.19	2.04	21	4.00	2.53	25
Int. Tech	3.80	2.19	30	4.43	1.86	21	3.87	2.03	24
Int. Science	3.17	2.14	29	3.19	2.36	21	2.25	1.42	24
Int. Leader	5.86	1.46	29	5.43	2.16	21	5.44	1.61	25
Int. Art	2.86	2.46	29	3.10	2.19	21	1.63	1.84	24
Int. Organize	5.00	1.81	29	5.62	1.83	21	4.75	1.85	24

Table 58. Mean Validity Ratings of Attributes for Effort and Leadership: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	5.10	1.18	30	5.81	1.57	21	4.21	1.98	24
Reason	5.83	1.23	30	5.43	1.86	21	4.96	1.54	25
Number	3.79	1.80	29	4.05	1.83	21	3.12	2.01	25
Spatial	4.37	1.71	30	4.29	1.52	21	4.00	2.02	25
Closure	4.68	1.70	28	4.38	1.75	21	3.76	2.39	25
Info Process	5.47	1.22	30	5.33	1.85	21	4.96	1.65	25
PSA	5.41	1.30	29	4.76	2.02	21	4.40	1.98	25
Memory	6.20	1.30	30	6.05	1.83	21	5.84	1.40	25
Mech Comp	4.62	1.61	29	5.48	1.81	21	3.58	2.32	24
Eye-Limb Coord	4.23	2.03	30	5.24	1.92	21	3.40	1.87	25
Precision	4.20	1.92	30	4.52	1.75	21	3.60	2.24	25
Movement Judge	3.55	1.90	29	4.10	2.02	21	2.72	2.11	25
Dexterity	3.93	2.13	30	4.43	2.09	21	2.92	2.31	25
Strength	5.33	1.60	30	5.67	1.39	21	3.76	1.98	25
Endurance	6.13	1.55	30	6.00	1.84	21	4.84	1.77	25
Balance	4.37	2.06	30	4.52	2.20	21	3.28	1.97	25
Athletics	5.10	2.09	29	4.95	1.99	21	3.64	2.25	25
Work Orient	6.50	1.63	30	6.81	1.47	21	5.16	1.86	25
Sociability	5.67	1.75	30	6.19	1.47	21	3.80	2.18	25
Cooperation	6.45	1.35	29	6.33	1.39	21	4.56	2.36	25
Energy	6.21	1.47	29	6.43	1.33	21	5.28	1.90	25
Conscientious	6.24	1.57	29	6.29	1.85	21	5.52	1.45	25
Dominance	6.55	1.35	29	6.52	1.69	21	5.52	2.02	25
Int. Tools	3.76	1.77	29	3.81	2.36	21	2.36	2.23	25
Int. Rugged	5.30	1.73	30	5.14	2.22	21	3.68	2.15	25
Int. Protect	5.20	1.63	30	5.62	1.72	21	3.32	2.06	25
Int. Tech	3.50	2.19	30	4.19	2.16	21	3.04	1.73	24
Int. Science	3.38	1.99	29	3.29	2.26	21	1.96	1.71	24
Int. Leader	6.69	1.37	29	7.00	1.22	21	6.00	1.50	25
Int. Art	2.86	2.57	29	3.29	2.10	21	1.46	1.56	24
Int. Organize	5.34	1.54	29	6.00	1.97	21	4.79	2.00	24

Table 59. Mean Validity Ratings of Attributes for Personal Discipline: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	3.53	2.37	30	3.67	2.42	21	3.29	2.22	24
Reason	4.50	1.70	30	4.86	1.77	21	3.60	1.91	25
Number	2.90	2.06	29	3.29	2.47	21	2.20	2.00	25
Spatial	3.30	2.05	30	3.48	2.04	21	2.80	2.22	25
Closure	2.96	1.91	28	3.43	2.06	21	2.72	2.21	25
Info Process	4.23	1.85	30	4.33	1.96	21	3.52	2.08	25
PSA	3.93	1.75	29	4.29	2.08	21	3.52	2.26	25
Memory	4.50	2.03	30	5.14	1.88	21	4.48	2.26	25
Mech Comp	3.66	2.13	29	3.57	2.68	21	2.25	2.09	24
Eye-Limb Coord	3.60	2.01	30	4.24	2.49	21	2.68	2.10	25
Precision	3.97	2.25	30	4.38	2.36	21	2.96	2.32	25
Movement Judge	3.10	1.95	29	3.10	2.30	21	1.84	1.72	25
Dexterity	3.37	2.09	30	3.52	2.64	21	2.60	2.18	25
Strength	5.27	1.89	30	4.10	2.51	21	3.48	2.28	25
Endurance	5.63	1.61	30	5.33	2.20	21	4.28	1.88	25
Balance	4.23	2.13	30	4.10	2.76	21	3.04	2.09	25
Athletics	5.17	2.09	29	4.90	1.95	21	3.56	2.08	25
Work Orient	6.40	1.75	30	6.38	1.40	21	4.80	2.22	25
Sociability	4.33	2.14	30	4.52	2.18	21	2.84	2.06	25
Cooperation	5.76	1.88	29	4.76	2.62	21	4.16	2.44	25
Energy	5.55	1.80	29	5.24	2.10	21	4.48	1.96	25
Conscientious	6.21	1.86	29	6.38	1.47	21	4.84	2.17	25
Dominance	5.41	1.72	29	5.38	2.38	21	5.24	1.88	25
Int. Tools	3.28	2.10	29	3.67	2.76	21	1.84	1.89	25
Int. Rugged	4.33	2.19	30	4.24	2.30	21	3.00	2.00	25
Int. Protect	4.47	2.18	30	4.90	2.72	21	2.71	1.63	24
Int. Tech	3.23	2.28	30	3.43	2.25	21	2.21	1.84	24
Int. Science	3.48	2.35	29	2.86	2.39	21	2.08	1.74	24
Int. Leader	6.00	1.71	29	5.95	2.09	21	4.88	1.83	25
Int. Art	2.83	2.45	29	3.24	2.07	21	1.79	1.61	24
Int. Organize	5.59	1.99	29	5.90	1.64	21	4.58	1.98	24

Table 60. Mean Validity Ratings of Attributes for Physical Fitness/Military Bearing: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	2.90	2.50	30	3.48	2.52	21	2.08	1.38	24
Reason	3.23	1.92	30	3.86	2.31	21	2.04	1.65	25
Number	2.21	1.93	29	2.81	2.32	21	1.36	1.75	25
Spatial	2.37	2.14	30	2.48	1.91	21	1.88	2.22	25
Closure	1.64	1.79	28	3.14	2.78	21	1.24	1.59	25
Info Process	2.60	1.77	30	3.67	2.31	21	2.32	1.77	25
PSA	2.90	2.32	29	3.52	2.16	21	1.88	1.79	25
Memory	3.03	2.14	30	3.90	2.55	21	2.48	2.08	25
Mech Comp	2.55	1.94	29	3.38	2.65	21	1.54	1.96	24
Eye-Limb Coord	5.00	2.10	30	5.19	2.02	21	3.88	2.60	25
Precision	4.07	2.27	30	4.38	2.42	21	3.28	2.88	25
Movement Judge	3.03	2.38	29	3.62	2.46	21	2.44	2.36	25
Dexterity	3.90	2.29	30	3.81	2.36	21	3.80	2.38	25
Strength	6.77	1.52	30	6.48	1.69	21	5.96	2.15	25
Endurance	6.87	1.28	30	6.76	1.61	21	6.48	1.45	25
Balance	5.30	1.70	30	5.57	1.96	21	5.04	2.21	25
Athletics	6.45	1.33	29	6.71	1.45	21	5.04	2.65	25
*Work Orient	5.67	1.77	30	6.10	1.64	21	3.52	2.37	25
Sociability	3.53	2.49	30	3.52	2.50	21	1.48	1.56	25
Cooperation	3.83	2.41	29	3.43	2.48	21	1.68	1.68	25
Energy	5.90	1.76	29	6.05	1.75	21	4.68	2.19	25
Conscientious	5.00	2.07	29	5.14	2.17	21	3.00	2.58	25
Dominance	5.07	1.77	29	4.95	2.92	21	3.20	2.22	25
Int. Tools	2.97	2.03	29	2.95	2.67	21	1.24	1.59	25
Int. Rugged	4.93	2.48	30	5.24	1.97	21	4.12	2.79	25
Int. Protect	3.67	2.26	30	3.38	2.44	21	2.52	2.10	25
Int. Tech	1.83	2.05	30	3.00	2.35	21	1.79	1.82	24
Int. Science	1.76	1.86	29	2.38	2.18	21	1.04	1.37	24
Int. Leader	5.24	2.17	29	4.95	2.42	21	3.20	2.48	25
Int. Art	1.72	1.96	29	2.86	2.29	21	1.29	1.68	24
*Int. Organize	3.31	2.35	29	5.29	2.24	21	2.58	1.91	24

Table 61. Mean Validity Rankings of Attributes: 11B, 63B, and 71L NCOs and Officers Combined

	11B			63B			71L		
Variable	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
Verbal	11.83	7.67	29	14.06	9.15	16	3.65	3.05	23
Reason	7.48	5.21	29	7.00	6.26	16	6.43	4.60	23
Number	21.38	5.43	29	15.06	6.64	16	14.30	7.62	23
Spatial	20.55	6.96	29	16.88	6.72	16	17.78	7.42	23
Closure	21.38	5.79	29	18.50	7.18	16	11.70	7.59	23
Info Process	9.24	6.19	29	10.19	6.42	16	7.91	6.17	23
PSA	15.55	7.31	29	12.19	6.86	16	10.09	7.32	23
Memory	8.76	5.15	29	10.87	7.45	16	8.00	4.67	23
Mech Comp	16.83	7.33	29	8.31	10.16	16	22.83	6.86	23
Eye-Limb Coord	14.66	5.94	29	13.25	7.33	16	12.61	6.95	23
Precision	16.76	6.24	29	13.25	8.38	16	13.70	6.49	23
Movement Judge	15.66	5.73	29	17.25	7.26	16	20.30	5.70	23
Dexterity	15.83	5.59	29	12.69	7.90	16	6.30	5.52	23
Strength	5.69	4.77	29	12.06	9.25	16	20.35	5.31	23
Endurance	3.79	2.82	29	14.69	8.06	16	19.26	5.75	23
Balance	15.34	5.76	29	15.50	6.22	16	22.52	3.98	23
Athletics	17.52	8.28	29	23.13	6.82	16	24.43	4.72	23
Work Orient	11.62	9.19	29	13.50	6.44	16	9.00	4.40	23
Sociability	16.97	6.96	29	22.38	8.29	16	13.22	6.86	23
Cooperation	12.93	7.37	29	15.63	7.10	16	10.30	6.26	23
Energy	10.76	6.23	29	18.12	7.68	16	12.17	4.48	23
Conscientious	14.79	8.10	29	14.94	6.21	16	10.70	5.69	23
Dominance	13.69	7.33	29	16.87	7.54	16	15.83	6.18	23
Int. Tools	23.69	5.78	29	10.44	10.20	16	23.70	6.64	23
Int. Rugged	17.45	9.96	29	22.81	6.60	16	27.70	2.70	23
Int. Protect	24.66	5.83	29	24.37	7.58	16	25.43	4.13	23
Int. Tech	24.72	6.65	29	12.44	7.99	16	21.35	8.30	23
Int. Science	27.66	3.07	29	23.50	9.70	16	27.22	3.54	23
Int. Leader	10.45	7.59	29	18.19	7.50	16	17.05	4.77	22
Int. Art	30.14	1.96	29	26.56	9.05	16	28.13	3.18	23
Int. Organize	18.24	9.30	29	17.06	8.74	16	12.65	7.76	23

APPENDIX A

DEMOGRAPHIC INFORMATION OF SMEs

Appendix A

Sample Demographic Information

MOS	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
11B	60	44.8	60	44.8
63B	35	26.1	95	70.9
71L	39	29.1	134	100.0

RANK	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
NCO	77	57.5	77	57.5
OFFICER	46	34.3	123	91.8
CIVILIAN	11	8.2	134	100.0

COMMAND	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
FORSCOM	66	49.3	66	49.3
DOTD	68	50.7	134	100.0

HGRADE	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
.	3	.	.	.
E4	1	0.8	1	0.8
E5	11	8.4	12	9.2
E6	20	15.3	32	24.4
E7	38	29.0	70	53.4
E8	12	9.2	82	62.6
E9	1	0.8	83	63.4
1ST LT	4	3.1	87	66.4
2ND LT	20	15.3	107	81.7
CAPTAIN	20	15.3	127	96.9
MAJOR	3	2.3	130	99.2
LT COL	1	0.8	131	100.0

SITE	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
ORD	28	20.9	28	20.9
BENJAMIN HARRISON	21	15.7	49	36.6
BENNING	26	19.4	75	56.0
ABERDEEN	21	15.7	96	71.6
STEWART	38	28.4	134	100.0

Sample Demographic Information

TABLE OF HGRADE BY MOS

HGRADE / MOS FREQUENCY PERCENT ROW PCT COL PCT	11B	63B	71L	TOTAL
E4	1 0.76 100.00 1.69	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76
E5	9 6.87 81.82 15.25	2 1.53 18.18 6.06	0 0.00 0.00 0.00	11 8.40
E6	9 6.87 45.00 15.25	9 6.87 45.00 27.27	2 1.53 10.00 5.13	20 15.27
E7	17 12.98 44.74 28.81	10 7.63 26.32 30.30	11 8.40 28.95 28.21	38 29.01
E8	4 3.05 33.33 6.78	1 0.76 8.33 3.03	7 5.34 58.33 17.95	12 9.16
E9	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76 100.00 2.56	1 0.76
TOTAL	59 45.04	33 25.19	39 29.77	131 100.00

(CONTINUED)

Sample Demographic Information

TABLE OF HGRADE BY MOS

HGRADE MOS

FREQUENCY PERCENT ROW PCT COL PCT	11B	63B	71L	TOTAL
1ST LT	3 2.29 75.00 5.08	0 0.00 0.00 0.00	1 0.76 25.00 2.56	4 3.05
2ND LT	11 8.40 55.00 18.64	5 3.82 25.00 15.15	4 3.05 20.00 10.26	20 15.27
CAPTAIN	4 3.05 20.00 6.78	5 3.82 25.00 15.15	11 8.40 55.00 28.21	20 15.27
MAJOR	0 0.00 0.00 0.00	1 0.76 33.33 3.03	2 1.53 66.67 5.13	3 2.29
LT COL	1 0.76 100.00 1.69	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76
TOTAL	59 45.04	33 25.19	39 29.77	131 100.00

FREQUENCY MISSING = 3

Sample Demographic Information

TABLE OF RANK BY COMMAND

RANK COMMAND

FREQUENCY PERCENT ROW PCT COL PCT	FORSCOM	DOTD	TOTAL
NCO	35 26.12 45.45 53.03	42 31.34 54.55 61.76	77 57.46
OFFICER	31 23.13 67.39 46.97	15 11.19 32.61 22.06	46 34.33
CIVILIAN	0 0.00 0.00 0.00	11 8.21 100.00 16.18	11 8.21
TOTAL	66 49.25	68 50.75	134 100.00

Sample Demographic Information

TABLE OF HGRADE BY COMMAND

HGRADE (FREQUENCY / PERCENT / ROW PCT / COL PCT)	COMMAND: FORSCOM	DOTD	TOTAL
E4	1 0.76 100.00 1.52	0 0.00 0.00 0.00	1 0.76
E5	10 7.63 90.91 15.15	1 0.76 9.09 1.54	11 8.40
E6	5 3.82 25.00 7.58	15 11.45 75.00 23.08	20 15.27
E7	18 13.74 47.37 27.27	20 15.27 52.63 30.77	38 29.01
E8	2 1.53 16.67 3.03	10 7.63 83.33 15.38	12 9.16
E9	0 0.00 0.00 0.00	1 0.76 100.00 1.54	1 0.76
TOTAL	66 50.38	65 49.62	131 100.00

(CONTINUED)

Sample Demographic Information

TABLE OF HGRADE BY COMMAND

HGRADE COMMAND

FREQUENCY PERCENT ROW PCT COL PCT	FORSCOM	DOTD	TOTAL
1ST LT	4 3.05 100.00 6.06	0 0.00 0.00 0.00	4 3.05
2ND LT	14 10.69 70.00 21.21	6 4.58 30.00 9.23	20 15.27
CAPTAIN	12 9.16 60.00 18.18	8 6.11 40.00 12.31	20 15.27
MAJOR	0 0.00 0.00 0.00	3 2.29 100.00 4.62	3 2.29
LT COL	0 0.00 0.00 0.00	1 0.76 100.00 1.54	1 0.76
TOTAL	66 50.38	65 49.62	131 100.00

FREQUENCY MISSING = 3

Sample Demographic Information

TABLE OF HGRADE BY RANK

HGRADE RANK

FREQUENCY PERCENT ROW PCT COL PCT	NCO	OFFICER	CIVILIAN	TOTAL
E4	1 0.76 100.00 1.30	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76
E5	11 8.40 100.00 14.29	0 0.00 0.00 0.00	0 0.00 0.00 0.00	11 8.40
E6	20 15.27 100.00 25.97	0 0.00 0.00 0.00	0 0.00 0.00 0.00	20 15.27
E7	37 28.24 97.37 48.05	1 0.76 2.63 2.17	0 0.00 0.00 0.00	38 29.01
E8	7 5.34 58.33 9.09	0 0.00 0.00 0.00	5 3.82 41.67 62.50	12 9.16
E9	1 0.76 100.00 1.30	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76
TOTAL	77 58.78	46 35.11	8 6.11	131 100.00

(CONTINUED)

Sample Demographic Information

TABLE OF HGRADE BY RANK

HGRADE RANK

FREQUENCY / PERCENT / ROW PCT / COL PCT	NCO	OFFICER	CIVILIAN	TOTAL
1ST LT	0 0.00 0.00 0.00	4 3.05 100.00 8.70	0 0.00 0.00 0.00	4 3.05
2ND LT	0 0.00 0.00 0.00	18 13.74 90.00 39.13	2 1.53 10.00 25.00	20 15.27
CAPTAIN	0 0.00 0.00 0.00	20 15.27 100.00 43.48	0 0.00 0.00 0.00	20 15.27
MAJOR	0 0.00 0.00 0.00	3 2.29 100.00 6.52	0 0.00 0.00 0.00	3 2.29
LT COL	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.76 100.00 12.50	1 0.76
TOTAL	77 58.78	46 35.11	8 6.11	131 100.00

FREQUENCY MISSING = 3

APPENDIX B

TASK CATEGORY INSTRUMENT

Name: ____________________

Rank: ____________________

MOS: ____________________

MOS TASK CATEGORY QUESTIONNAIRE

Listed below are 98 task categories. For each task category, we would like you to make the following four ratings:

1. Is the task a **PART OF THE JOB** for E3 and E4 soldiers in the MOS you are rating?

   _N_ = No (skip ratings 2 - 4 and go to the next task)

   _Y_ = Yes (continue with ratings 2 - 4)

2. How **IMPORTANT** or critical is the task for successful performance for E3 and E4 soldiers in the MOS?

   _5_ = Extremely high importance for E3 and E4 soldiers. The task is very central to the job. (Only the most critical tasks should get this rating.)

   _4_ = High importance. The task is central to the job, but it is not one of the most critical tasks.

   _3_ = Moderate importance. The task is part of the job, but it is not central to the job.

   _2_ = Low importance. The task is part of the job, but there are many tasks that are more important.

   _1_ = Very minor importance. The task is part of the job, but it is not at all important for successful job performance.

3. How **FREQUENTLY** is the task performed by E3 and E4 soldiers in the MOS?

   _3_ = Very frequently. The task is performed much more frequently than most other tasks.

   _2_ = Frequently. The task is performed about as often as other tasks.

   _1_ = Occasionally or infrequently. The task is performed much less frequently than most other tasks.

4. How **DIFFICULT** and demanding is it for E3 and E4 soldiers in the MOS to perform this task well?

- 3 = Very difficult. It takes considerable skill and motivation to perform this task well. Few soldiers can perform this task.
- 2 = Moderately difficult. It requires some skill and motivation to perform this task well, but most soldiers are eventually able to learn the task.
- 1 = Routine. The task is relatively easy and requires little skill. Almost all soldiers learn the task very quickly and are able to perform the task with little difficulty.

	Part of Job?	Importance	Frequency	Difficulty

I. Maintenance

A. Mechanical Systems Maintenance

1. __Perform operator maintenance checks and services__ -- follow directions in Operator's Manual; conduct before, during, after, and weekly operator checks and services on vehicles, trailers, generators, construction equipment, or other kinds of mechanical apparatus. ____ ____ ____ ____

2. __Perform operator checks and services on weapons__ -- check, disassemble, assemble, clean, lubricate, and adjust weapons, including pistols, rifles, machineguns, hand grenades, and breechblocks. ____ ____ ____ ____

3. __Inspect mechanical systems__ -- measure, use test equipment and manuals, and observe mechanical equipment (e.g., engines, transmissions, brakes, hydraulics, refrigeration systems) to detect problems and malfunctions. ____ ____ ____ ____

4. __Repair weapons__ -- after the cause of a problem in a weapon has been found, fix it using the appropriate tools and necessary replacement parts by following directions in the weapon's technical manual. ____ ____ ____ ____

5. __Repair mechanical systems__ -- after the cause of a problem in a mechanical part has been found, fix it using the appropriate tools (e.g., wrenches, screwdrivers, gauges, hammers, soldering equipment) and necessary replacement parts by following directions in the equipment's technical manual. ____ ____ ____ ____

Part of Job? | Importance | Frequency | Difficulty

6. <u>Troubleshoot mechanical systems</u> -- find the cause of malfunctions in mechanical parts and equipment using technical manuals, tools, and test equipment (e.g., calipers, gauges, torque wrenches, pressure gauges).

7. <u>Troubleshoot weapons</u> -- find the cause of malfunctions in weapons using technical manuals, tools, and test equipment.

B. Electrical and Electronic Systems Maintenance

8. <u>Install electronic components</u> -- connect electronic and communications equipment (e.g., radios, antennas, telephones, teletypewriters, radar, power supplies) and check system for operation.

9. <u>Inspect electrical systems</u> -- measure, use test equipment and manuals, and observe electrical systems (e.g., generators, wiring harnesses, switches, relays, circuit breakers) to detect problems and malfunctions.

10. <u>Inspect electronic systems</u> -- measure, use test equipment and manuals, and observe electronic systems (e.g., communications equipment, radar, missile and tank computer ballistics) to detect problems and malfunctions.

11. <u>Repair electrical systems</u> -- after the cause of an electrical problem has been found, fix it with the appropriate tools (e.g., wire strippers, pliers, soldering irons) and necessary replacement parts by following directions in the equipment's technical manual.

	Part of Job?	Importance	Frequency	Difficulty
12. <u>Repair electronic components</u> -- After the cause of an electronics problem has been identified, fix it with appropriate tools (e.g., tests sets, screwdrivers, pliers, soldering guns) and necessary replacement parts by following directions in the equipment's technical manual.	____	____	____	____
13. <u>Troubleshoot electrical systems</u> -- find the cause of malfunctions in electrical parts and equipment using technical manuals, tools, and test equipment (e.g., multimeters, test lights).	____	____	____	____
14. <u>Troubleshoot electronic components</u> -- find the cause of malfunctions in electronic equipment and components using technical manuals, tools and test equipment (e.g., oscilloscopes, specialized test sets).	____	____	____	____

II. GENERAL OPERATIONS
C. Pack and Load

	Part of Job?	Importance	Frequency	Difficulty
15. <u>Pack and load materials</u> -- load and lash materials onto transport vehicles (land, sea, or air) to secure and protect from damage or loss during shipment.	____	____	____	____
16. <u>Prepare parachutes</u> -- inspect cargo and personnel parachutes, repair or replace faulty components, and pack parachutes for air drop.	____	____	____	____
17. <u>Prepare equipment and supplies for air drop</u> -- build or assemble platforms, cushions, and riggings for parachuting supplies, equipment, and vehicles.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
D. Vehicle and Equipment Operations				
18. <u>Operate power excavating equipment</u> -- use air hammers and drills, paving breakers, grinders, backfill tampers, or other hand operated power equipment in building concrete, stone, or other structures (e.g., roads, fortifications, buildings).	____	____	____	____
19. <u>Operate wheeled vehicles</u> -- drive wheeled vehicles over roads and cross-country in response to mission, terrain, and traffic regulations.	____	____	____	____
20. <u>Operate track vehicles</u> -- drive track vehicles (e.g., tank, APC, BFV) in response to mission, terrain, and traffic controls.	____	____	____	____
21. <u>Operate boats</u> -- drive boats and rafts.	____	____	____	____
22. <u>Operate lifting, loading, and grading equipment</u> -- operate fork lifts, cranes, back-hoes, graders, and other heavy equipment to load, unload, or move heavy equipment, supplies, construction materials (e.g., culvert pipe, building and bridge parts), or terrain (e.g., earth, rocks, trees).	____	____	____	____
E. Construct/Assemble				
23. <u>Paint</u> -- prepare surfaces (clean, remove old paint, sand) and apply paint with brush, roller, or spray.	____	____	____	____
24. <u>Install wire and cables</u> -- string or lay, and connect electrical wire or communications cables.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
25. Repair plastic and fiberglass -- fix plastic or fiberglass parts and structures by cutting, sawing, drilling, sanding, filling, gluing, and painting.	____	____	____	____
26. Repair metal -- fix metal structures or parts by bending, cutting, drilling, welding, hammering, grinding, soldering, and painting.	____	____	____	____
27. Assemble steel structures -- erect bridges, antennas, and other steel structures. May require the assistance of others and use of heavy equipment.	____	____	____	____
28. Install pipe assemblies -- place, connect, and test pipe assemblies and fixtures (e.g., plumbing, POL pipelines and pumps).	____	____	____	____
29. Construct wooden buildings and other structures -- measure, saw, nail, plane to frame, sheath, and roof buildings, or erect trestles, bridges, and piers.	____	____	____	____
30. Construct masonry buildings and structures -- measure, lay brick or concrete blocks, or build forms and pour concrete to construct walls, columns, field fortifications, and other concrete or masonry structures.	____	____	____	____
F. Technical Procedures				
31. Operate gas and electric powered equipment -- operate electric generators, air compressors, smoke generators, quarry machines, mobile washing machines, water pumps, etc., to produce power or process materials.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
32. <u>Select, layout and clean medical or dental equipment and supplies</u> -- prepare treatment areas for use by following prescribed procedures for laying-out instruments and equipment; clean equipment and area for future use.	____	____	____	____
33. <u>Use audiovisual equipment</u> -- use cameras and videotape to record sights and sounds for intelligence analysis, training, or documentation.	____	____	____	____
34. <u>Reproduce printed material</u> -- operate duplicating machines, offset presses, and similar equipment to reproduce printed materials; collate and bind materials using various types of bindery equipment.	____	____	____	____
35. <u>Operate electronic equipment</u> -- set and adjust the controls to operate electronic equipment (e.g., radio, computer hardware, missile ballistics controls).	____	____	____	____
36. <u>Operate radar</u> -- operate radar equipment and interpret signals.	____	____	____	____
37. <u>Operate computer hardware</u> -- operate computer hardware such as tape and disk drives, optical scanners, terminals, and other input/output devices.	____	____	____	____
38. <u>Cook</u> -- prepare food and beverages according to recipes and meal plans (measure, mix, bake); inspect fresh food and staples for freshness; clean equipment and work area.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
39. <u>Perform medical laboratory procedures</u> -- conduct various types of blood tests, urinalysis, cultures, etc.	____	____	____	____
40. <u>Conduct land surveys</u> -- survey terrain to determine elevations, azimuths, and distances of terrain features; record information.	____	____	____	____
41. <u>Provide medical or dental treatment</u> -- give medical attention to soldiers in the field or in medical or dental clinics, or give veterinary treatment to animals (e.g., administer injections, take blood pressure, change sterile dressings); does not include first aid.	____	____	____	____
G. Make Technical Drawings				
42. <u>Sketch maps, overlays, or range cards</u> -- use standard symbols to make sketches of terrain, including locations of buildings and other objects, targets, avenues of approach, and maneuver areas.	____	____	____	____
43. <u>Produce technical drawings</u> -- use drafting and drawing equipment to make technical drawings and blueprints.	____	____	____	____
44. <u>Draw maps and overlays</u> -- use drafting, graphics, and related techniques to draw and revise maps from aerial photographs.	____	____	____	____
45. <u>Draw illustrations</u> -- using pen, pencil, paint, or other media, make free hand drawings and illustrations.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty

III. ADMINISTRATIVE

H. Clerical

46. Type -- type information using a typewriter, teletypewriter, keypunch, or computer terminal. ____ ____ ____ ____

47. Prepare technical forms and documents -- follow standardized procedures to prepare or complete forms and documents (e.g., personnel records, legal briefs). ____ ____ ____ ____

48. Record, file, and dispatch information -- collect, copy, update, sort, index, file, and retrieve information (e.g., mail, training rosters, personnel statistics, supply inventories). ____ ____ ____ ____

49. Receive, store, and issue supplies, equipment, and other materials -- inspect materials and review paperwork when receiving materials; sort, transport, and store materials; issue or ship material to authorized personnel or units. ____ ____ ____ ____

I. Communication

50. Use hand and arm signals -- communicate messages and instructions using hand and arm signals. ____ ____ ____ ____

51. Read technical manuals, field manuals, regulations, and other publications -- use index and table of contents to find location of needed information; locate information; read instructions, diagrams, charts, and tables. ____ ____ ____ ____

52. Use maps -- read and interpret map symbols and identify terrain features in order to orient map to your position in the field; determine grid coordinates; determine directions; identify roads, towns, etc. ____ ____ ____ ____

	Part of Job?	Importance	Frequency	Difficulty
53. <u>Send and receive radio messages</u> -- use standardized radio codes and procedures to transmit and receive messages and other information.	____	____	____	____
54. <u>Give oral reports</u> -- use standard communication procedures to organize information (e.g., SALUTE, challenge and password, call for fire).	____	____	____	____
55. <u>Receive clients, patients, guests</u> -- schedule, greet, and give routine information to persons seeking medical, dental, legal, or counseling services.	____	____	____	____
56. <u>Give directions and instructions</u> -- verbally give information, instructions, or directions to others.	____	____	____	____
57. <u>Write documents and correspondence</u> -- draft letters, reports, memos, etc.; proofread and edit.	____	____	____	____
58. <u>Write and deliver presentations</u> -- prepare scripts for formal presentations, including radio and television broadcasts and briefings.	____	____	____	____
59. <u>Interview</u> -- gather information from clients, patients, witnesses, prisoners, or other persons.	____	____	____	____
60. <u>Provide counseling and other interpersonal interventions</u> -- conduct personal adjustment counseling with individuals and groups; use interpersonal relations skills to solve relationship problems.	____	____	____	____

J. Analyze Information	Part of Job?	Importance	Frequency	Difficulty
61. Decode data -- use coding systems and rules to decipher and interpret coded information (e.g., use CEOI, interpret symbols/signs).	____	____	____	____
62. Analyze electronic signals -- analyze electronics signals to detect threat transmitters and electronic countermeasures.	____	____	____	____
63. Analyze weather conditions -- determine weather conditions and analyze their effects on tactical operators.	____	____	____	____
64. Order equipment and supplies -- determine needs and requisition needed supplies, materials, and equipment.	____	____	____	____
65. Estimate time and cost of maintenance operations -- estimate equipment downtime and cost of repairs, including parts and labor.	____	____	____	____
66. Plan placement or use of tactical equipment -- using maps and on-site inspection, identity positions and areas to be used for cover and concealment and to place weapons, fortifications, mines, and detectors.	____	____	____	____
67. Translate foreign languages -- translate written or spoken foreign language communications.	____	____	____	____
68. Analyze intelligence data -- determine importance and reliability of information; use information to determine identity, capabilities, disposition, and movement of enemy forces.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty

K. Applied Math and Data Processing

69. <u>Control money</u> -- keep accounting records; disperse and collect money and money orders. ____ ____ ____ ____

70. <u>Determine firing data for indirect fire weapons</u> -- using maps, firing charts, and targeting and ballistics information, determine elevation and azimuth needed for engaging targets. ____ ____ ____ ____

71. <u>Compute statistics or other mathematical calculations</u> -- select formulas and make mathematical calculations, with or without using calculators or computers; report results. ____ ____ ____ ____

72. <u>Provide programming and data processing support for computer operations</u> --analyze data processing needs; select or prepare, edit, test, and run computer programs; document process and results. ____ ____ ____ ____

L. Control Air Traffic

73. <u>Control air traffic</u> -- coordinate departing, en route, arriving, and holding aircraft by monitoring radar equipment, communicating with aircraft and other air traffic control units. ____ ____ ____ ____

IV. COMBAT

M. Individual Combat

74. <u>Use hand grenades</u> -- identify, inspect, arm, throw, and secure hand grenades. ____ ____ ____ ____

	Part of Job?	Importance	Frequency	Difficulty
75. <u>Protect against NBC hazards</u> -- use protective clothing, masks, and decontamination equipment to protect self, others, equipment, and supplies from nuclear, biological, and chemical hazards.	____	____	____	____
76. <u>Handle demolitions or mines</u> -- store, place, charge, discharge, and disarm explosives, demolition devices, or mines.	____	____	____	____
77. <u>Engage in hand-to-hand combat</u> -- use offensive and defensive maneuvers to overcome hostile individuals.	____	____	____	____
78. <u>Fire individual weapons</u> -- aim, track and fire individual weapons, such as rifles, pistols, machineguns, and LAW at designated targets; load, reduce a stoppage, and clear weapons.	____	____	____	____
79. <u>Control individuals and crowds</u> -- apprehend and search suspected criminals or enemy soldiers; guard prisoners; participate in riot control; perform guard duty.	____	____	____	____
80. <u>Customs and laws of war</u> -- use knowledge of Geneva convention and military SOP concerning treatment of enemy personnel, engagement of the enemy, conduct of military protocol and ceremony, guard duty, and physical readiness.	____	____	____	____
81. <u>Navigate</u> -- during the day or night, with or without a map, locate positions and move from point to point in response to terrain features (e.g., for cover or concealment), battle conditions, and mission.	____	____	____	____

	Part of Job?	Importance	Frequency	Difficulty
82. <u>Survive in the field</u> -- react to direct or indirect fire; prepare individual battle positions; camouflage self and equipment; observe security procedures.	____	____	____	____
N. Crew-served Weapons				
83. <u>Load and unload field artillery or tank guns</u> -- operate breech controls and handle ammunition (stow and load) to prepare guns for firing; unload or extract unused rounds or misfires.	____	____	____	____
84. <u>Fire heavy direct fire weapons</u> (e.g., tank main guns, TOW missile, IFV cannon) -- using weapon sights, manipulate weapon controls to aim, track, and fire on targets.	____	____	____	____
85. <u>Prepare heavy weapons for tactical use</u> -- position and prepare for firing heavy tactical weapons, such as missiles, field artillery, and anti-aircraft systems.	____	____	____	____
86. <u>Place and camouflage tactical equipment and materials in the field</u> -- place mines, detectors, chemicals, and camouflage materials into position in the battlefield.	____	____	____	____
87. <u>Fire indirect fire weapons</u> (e.g., field artillery) -- lay weapon by adjusting azimuth and elevation controls in response to fire commands.	____	____	____	____
O. Give First Aid				
88. <u>Give first aid</u> -- carry out first aid procedures (e.g., CPR, put on field dressing, prevent shock).	____	____	____	____

Part of Job?	Importance	Frequency	Difficulty

P. Identify Targets

89. **Detect and identify targets** -- with or without optical devices (e.g., night sights, weapon sights, binoculars), locate possible targets, and identify type (e.g., troops, tanks, aircraft) and nomenclature. _____ _____ _____ _____

IV. Supervision

90. Plan, organize, monitor -- assign work tasks, supervise performance of tasks, conduct inspections, and monitor equipment condition and supplies. _____ _____ _____ _____

91. Clarify roles, provide feedback - monitor performance and counsel subordinates. _____ _____ _____ _____

92. Provide information - pass on information concerning mission and requirements to subordinates. _____ _____ _____ _____

93. Recognize, reward - provide formal and informal rewards and recognition for good performance; recommend soldiers for promotion or awards. _____ _____ _____ _____

94. Support - listen to subordinates' personal problems; counsel, assist, or arrange assistance, as appropriate. _____ _____ _____ _____

95. Train, develop - plan and conduct individual and team training, provide career counseling, and provide opportunities for leadership. _____ _____ _____ _____

96. Discipline, punish - provide formal or informal disciplinary measures to subordinates. _____ _____ _____ _____

	Part of Job?	Importance	Frequency	Difficulty
97. <u>Act as model</u> - set the example for subordinates.	____	____	____	____
98. <u>Conduct tactical operations</u> - supervise and direct activities of the unit during tactical operations.	____	____	____	____

Name ________________________________

TASK CATEGORY RATING EXERCISE
EVALUATION SHEET

Task Category Definitions

1. How clear were the task category definitions?

1	2	3	4	5	6	7
Not at all clear			Neither clear nor unclear			Very clear

2. How well did you understand the terms used in the task category definitions?

1	2	3	4	5	6	7
There were many terms I did not understand			There were some terms I did not understand			There were no terms I did not understand

3. Which task category was most difficult to judge? ____________

4. Which task category was least difficult to judge? ____________

5. What percentage of your job was covered by these task categories?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

If you answered less than 100%, what task categories should be added?

____

____

____

____

____

____

Task Category Ratings

1. How clear were the instructions for making the ratings?

1	2	3	4	5	6	7
Not at all clear			Neither clear nor unclear			Very clear

2. Were the ratings easy or difficult to make?

1	2	3	4	5	6	7
Very easy			Neither easy nor difficult			Very difficult

3. How confident did you feel about the accuracy of your ratings?

1	2	3	4	5	6	7
Not at all confident			Somewhat confident			Very confident

General Evaluation

1. Which of the following would you change (check all that apply):

   _____ Task Category Definitions

   _____ Task Category Rating Instructions and Procedures

2. What changes would you make?

____

____

____

____

____

____

MATCHING EXERCISE INSTRUCTIONS: MOS TASKS AND TASK CATEGORIES

In this portion of the workshop, we are interested in finding out how well our list of 98 task categories can be matched up with specific tasks from the Soldier's Manual (SM) and from the Army Occupational Survey Program (AOSP).

We have reviewed the SM and the AOSP and have selected 30 tasks that cover the range of responsibilities for Skill Level 1 and Skill Level 2 soldiers. Following each task, there are three blank lines. In those blanks, we want you to list the numbers of up to 3 task categories that best describe that task. You can find the names and numbers of the task categories on the "MOS Task Category Questionnaire" that you completed during the previous exercise. This form provides the number, the name, and a brief description of each of the 98 task categories. Or, if you would prefer, the task category names and numbers also can be obtained from the attached handout, "List of Task Categories" (which does not include any descriptions of the categories).

As you review the task categories, first find the task category that best describes the task. Put the number of this task category in the first column, under "1st Category/Activity." Then find the number of the task category that provides the second best description of the task and mark this number under "2nd Category/Activity." Finally, do the same for the third task category.

You do not have to fill in all three blanks for all of the tasks. You may find that only one or two categories describe some tasks. Feel free to leave the remaining column(s) blank. If you find that none of the task categories describes a task, simply write "NONE" across the three blanks.

Also, you are not limited in the number of times that you can use one of the task categories. For example, if the same task category describes all of the tasks on the Task List, you can use that category for every task.

List of Task Activities

1. Perform operator maintenance checks and services
2. Perform operator checks and services on weapons
3. Inspect mechanical systems
4. Repair weapons
5. Repair mechanical systems
6. Troubleshoot mechanical systems
7. Troubleshoot weapons
8. Install electronic components
9. Inspect electrical systems
10. Inspect electronic systems
11. Repair electrical systems
12. Repair electronic components
13. Troubleshoot electrical systems
14. Troubleshoot electronic components
15. Pack and load materials
16. Prepare parachutes
17. Prepare equipment and supplies for air drop
18. Operate power excavating equipment
19. Operate wheeled vehicles
20. Operate track vehicles
21. Operate boats
22. Operate lifting, loading, and grading equipment
23. Paint
24. Install wire and cables
25. Repair plastic and fiberglass
26. Repair metal
27. Assemble steel structures
28. Install pipe assemblies
29. Construct wooden buildings and other structures
30. Construct masonry buildings and structures
31. Operate gas and electric powered equipment
32. Select, layout and clean medical or dental equipment
33. Use audiovisual equipment
34. Reproduce printed material
35. Operate electronic equipment
36. Operate radar
37. Operate computer hardware
38. Cook
39. Perform medical laboratory procedures
40. Conduct land surveys
41. Provide medical or dental treatment
42. Sketch maps, overlays, or range cards
43. Produce technical drawings
44. Draw maps and overlays
45. Draw illustrations
46. Type
47. Prepare technical forms and documents
48. Record, file, and dispatch information
49. Receive, store, and issue supplies, equipment, and other materials

50. Use hand and arm signals
51. Read technical manuals, field manuals, regulations, and other publications
52. Use maps
53. Send and receive radio messages
54. Give oral reports
55. Receive clients, patients, guests
56. Give directions and instructions
57. Write documents and correspondence
58. Write and deliver presentations
59. Interview
60. Provide counseling and other interpersonal interventions
61. Decode data
62. Analyze electronic signals
63. Analyze weather conditions
64. Order equipment and supplies
65. Estimate time and cost of maintenance operations
66. Plan placement or use of tactical equipment
67. Translate foreign languages
68. Analyze intelligence data
69. Control money
70. Determine firing data for indirect fire weapons
71. Compute statistics or other mathematical calculations
72. Provide programming and data processing support for computer operations
73. Control air traffic
74. Use hand grenades
75. Protect against NBC hazards
76. Handle demolitions or mines
77. Engage in hand-to-hand combat
78. Fire individual weapons
79. Control individuals and crowds
80. Customs and laws of war
81. Navigate
82. Survive in the field
83. Load and unload field artillery or tank guns
84. Fire heavy direct fire weapons (e.g., tank main guns, TOW missile, IFV cannon)
85. Prepare heavy weapons for tactical use
86. Place and camouflage tactical equipment and materials in the field
87. Fire indirect fire weapons (e.g., field artillery)
88. Give first aid
89. Detect and identify targets
90. Plan, organize, monitor
91. Clarify roles, provide feedback
92. Provide information
93. Recognize, reward
94. Support
95. Train, develop
96. Discipline, punish
97. Act as model
98. Conduct tactical operations

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
24.	Put on and wear protective clothing in accordance with established (MOPP) levels	____	____	____
25.	Put on a field or pressure dressing	____	____	____
26.	Perform expedient repairs	____	____	____
27.	Troubleshoot engine cooling system	____	____	____
28.	Troubleshoot engines	____	____	____
29.	Troubleshoot steering system	____	____	____
30.	Slave start disabled vehicle	____	____	____

NAME________________________________

TASK MATCHING EXERCISE

Task List for 71L
Form A

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
1.	Type a basic comment to a disposition form	____	____	____
2.	Type a military letter	____	____	____
3.	Type a joint messageform (DD Form 173/1)	____	____	____
4.	Prepare a requisition for publications and/or blank forms--AUTODIN (DA Form 4569)	____	____	____
5.	Type military orders	____	____	____
6.	Type a memorandum	____	____	____
7.	Dispatch outgoing dis-tribution	____	____	____
8.	Establish functional files	____	____	____
9.	File documents/correspon-dence	____	____	____
10.	Type a second or subsequent comment to a disposition form	____	____	____
11.	Type straight copy material	____	____	____
12.	Assemble correspondence	____	____	____
13.	Receipt/transfer classified material	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
14.	Safeguard "FOR OFFICIAL USE ONLY" (FOUO) material	____	____	____
15.	Perform operator maintenance on an M16A1 rifle, magazine, and ammunition	____	____	____
16.	Load, reduce a stoppage, and clear an M16A1 rifle	____	____	____
17.	Camouflage yourself and your individual equipment	____	____	____
18.	Practice noise, light, and litter discipline	____	____	____
19.	Determine the grid coordinates of a point on a military map using the military grid reference system	____	____	____
20.	Determine a magnetic azimuth using a compass	____	____	____
21.	Put on a field or pressure dressing	____	____	____
22.	Maintain an M17 series protective mask	____	____	____
23.	Put on and wear an M17 series protective mask with hood	____	____	____
24.	Put on and wear protective clothing in accordance with established mission-oriented protective posture (MOPP) levels	____	____	____
25.	Know your rights and obligations as a prisoner of war	____	____	____
26.	Administer nerve agent antidote to self (self-aid)	____	____	____

Task No.	Task Name	1st Task Category/ Activity	2nd Task Category/ Activity	3rd Task Category/ Activity
27.	Receive, maintain, and control office equipment	____	____	____
28.	Control expendable/non-expendable supplies	____	____	____

APPENDIX D

ATTRIBUTE INSTRUMENT

ATTRIBUTE DEFINITIONS

Verbal Ability

This is the ability to use and understand spoken and written language and to communicate with others. It involves "catching on" to what's happening and coming up with and understanding words and ideas.

An individual HIGH on VERBAL ABILITY understands the meaning of even difficult or unusual words; understands written material quickly; rapidly comes up with words or ideas; quickly identifies the relationship between words or ideas; and, clearly communicates ideas and information.

An individual who has AVERAGE VERBAL ABILITY understands the meaning of most common words but does not know very difficult words; understands most written material in a reasonable amount of time, but may not pick up subtle ideas or complex directions; can come up with appropriate words or ideas in reasonable time; can identify obvious relationships between words or ideas quickly but requires more time to grasp complicated or unusual relationships; and, communicates ideas and information moderately well.

An individual LOW on VERBAL ABILITY does not know the meaning of even common words or phrases; cannot understand most written material; has difficulty coming up with appropriate words or ideas; cannot identify the relationships between most words or ideas; and, has difficulty communicating ideas and information.

Reasoning

This is the ability to discover a rule or principle underlying the relationship between two or more objects and apply it in solving a problem. It also includes the ability to use logic in drawing conclusions from available information.

An individual HIGH on REASONING who is shown a set of objects can quickly identify the common theme or characteristic that those objects share, and can tell whether new objects do or do not share the same characteristic; can study a sequence of objects or figures and can readily determine what changes occur from the first to the second object, from the second to the third object, and so on, and can quickly tell what the next object in the series would look like; and, when given a set of facts and a set of conclusions, can quickly determine which conclusions are correct.

An individual with AVERAGE REASONING who is shown a set of objects may occasionally have trouble identifying the common theme or characteristic that those objects share, and so may not be able to tell whether new objects do or do not share the same characteristic; usually can determine what changes occur from one object to the next in a series of objects, but may occasionally have trouble; and sometimes may have difficulty deciding what the next object in the series would look like; and, when given a set of facts and a set of conclusions, may have difficulty determining which conclusions are correct.

An individual LOW on REASONING who is shown a set of objects often has a great deal of trouble identifying the common theme or characteristic that those objects share, and so often will not be able to tell whether new objects do or do not share the same characteristic; has trouble determining what changes occur from one object to the next in a series of objects, and, therefore, is not able to tell what the next object in the series would look like; and, when given a set of facts and a set of conclusions, cannot determine which conclusions are correct.

Number Ability

This is the ability to perform arithmetic calculations and manipulations quickly and correctly, and to solve mathematical word problems.

An individual HIGH on NUMBER ABILITY performs simple computations rapidly without error (i.e., addition, subtraction, multiplication, and division); can use formulas to quickly and accurately solve almost all number problems; and, can quickly and correctly solve even difficult mathematical word problems.

An individual who has AVERAGE NUMBER ABILITY performs simple computations with moderate speed and few errors (i.e., addition, subtraction, multiplication, and division); can use formulas to solve some number problems in reasonable time with minimal errors; and, can solve simple mathematical word problems, but may be unable to solve difficult word problems without help.

An individual LOW on NUMBER ABILITY has difficulty performing simple computations (i.e., addition, subtraction, multiplication, and division); cannot use formulas to solve number problems; and, cannot solve even simple mathematical word problems.

Spatial Ability

This is the ability to locate objects that are partially hidden, or to identify objects that have been turned backwards or upside down or have been rotated sideways. It includes the ability to mentally break a figure into pieces and come up with a new arrangement, and the ability to remember how an object was turned or configured and where it was located.

An individual HIGH on SPATIAL ABILITY can quickly find a simple form that has been hidden inside a complex pattern; nearly always can remember where an object was located and how it was turned or configured; can quickly and accurately identify a path through a complex maze; and, always can identify exactly how a complex object will look when it has been rearranged or turned upside down, backwards, or sideways.

An individual with AVERAGE SPATIAL ABILITY can find a simple form that has been hidden inside a complex pattern, though it may require some time; usually, but not always, can remember where an object was located and how it was turned or configured; can quickly and accurately identify a path through a simple maze, but will require a great deal of time and have difficulty identifying a path through a complex maze; and, usually can identify how a complex object will look when it has been rearranged or turned upside down, backwards, or sideways.

An individual LOW on SPATIAL ABILITY has great difficulty and may not be able to find a simple form that has been hidden inside a complex pattern; usually, but not always, cannot remember where an object was located or how it was turned or configured; has great difficulty identifying a path through even a simple maze; and, cannot identify how a complex object will look when it has been rearranged or turned upside down, backwards, or sideways.

Closure

This is the ability to identify objects or words, given only sketchy or partial information.

An individual HIGH on CLOSURE is able to read letters and words and recognize objects quickly, even if a large portion of the letter, word, or object has been hidden or erased.

An individual with AVERAGE CLOSURE is able to read most letters and words and recognize most objects when portions of the letter, word, or object have been hidden or erased, but may require a fair amount of time to recognize these items; and, may be unable to recognize the item if a large portion has been hidden or erased.

An individual LOW on CLOSURE has great difficulty recognizing letters, words, and objects that have been partially hidden or erased; and, may require a great deal of time to recognize items that are only slightly hidden or erased.

Mental Information Processing

This is the ability to react quickly to a stimulus. It also involves the ability to attend carefully to information that is being presented (e.g., in a conversation, on a computer screen), even when other events are occurring and competing for one's attention. Finally, it involves the ability to direct attention to and perform two tasks at the same time, when required.

The individual HIGH on MENTAL INFORMATION PROCESSING is able to react extremely quickly to sounds (e.g., bells, snapping twigs) and sights (lights, unusual shadows); can focus attention on one event or problem, and can block out distracting noises, sights, or events that are competing for attention; when required, is able to focus attention on and perform two tasks very well at the same time (e.g., tracking a blip closely on radar while doing computations on a calculator).

The individual with AVERAGE MENTAL INFORMATION PROCESSING ability is able to react fairly quickly to sounds (e.g., bells, snapping twigs) and sights (lights, unusual shadows); usually can focus attention on one event or problem, but may occasionally have trouble blocking out distracting noises, sights, or events that are competing for attention; when required, is able to focus attention on and perform two tasks at the same time, but performance on one or both of the tasks may be rather slow or inaccurate.

The individual LOW on MENTAL INFORMATION PROCESSING reacts rather slowly to sounds (e.g., bells, snapping twigs) and sights (lights, unusual shadows); has great difficulty focusing attention on one event or problem when distracting noises, sights, or events are competing for attention; has trouble focusing attention on and performing more than one task at a time, and generally will ignore one task completely while performing the other.

Perceptual Speed and Accuracy

This is the ability to notice details about things (letters, objects, numbers, symbols, or patterns) quickly and correctly. This involves rapidly noting changes and/or the way things differ or are alike.

An individual HIGH on PERCEPTUAL SPEED AND ACCURACY is able to see and process things quickly and accurately; can quickly perform complex perceptual tasks (e.g., compare several objects and eliminate all those that are not identical); and, is rarely distracted by irrelevant cues.

An individual who has AVERAGE PERCEPTUAL SPEED AND ABILITY can see and process things with moderate speed and few mistakes; can perform simple perceptual tasks well (e.g., compare two objects for similarity), but has difficulty with more complex tasks; and, is occasionally distracted by irrelevant cues.

An individual LOW on PERCEPTUAL SPEED AND ACCURACY takes a great deal of time to see and process things; makes many errors when performing simple perceptual tasks; and, is easily distracted by irrelevant cues.

Memory

This is the ability to recall previously learned material. It includes the ability to remember miscellaneous facts and trivia, as well as the ability to recall a series of related facts, principles, concepts, words, pictures, and ideas.

An individual HIGH on MEMORY is able to accurately recall information from conversations, books, or magazine articles with little or no prompting; and, is able to recall a great deal of "figural" information, such as pictures, maps, faces, and scenes from places the individual has visited.

An individual with AVERAGE MEMORY is able to recall quite a bit of information from conversations, books, or magazine articles, but may require some prompting and may omit some details; and, is able to recall most figural information, but may have trouble recalling certain features of maps, pictures, or faces.

An individual LOW on MEMORY is able to recall some information from conversations, books, or magazine articles, but often requires a great deal of prompting and omits significant details; has trouble recalling figural information; and, may not be able to recognize many maps, pictures, or faces that have been encountered previously.

Mechanical Comprehension

This is the ability to understand mechanical, shop, automotive, and electrical terms and knowledge, and the way machines, tools, and equipment operate.

An individual HIGH on MECHANICAL COMPREHENSION can quickly recognize and understand the relationship of physical forces and mechanical elements in practical situations (e.g., how levers and gears work, what causes high pressure to occur in the flow of water); is able to learn and figure out mechanical principles and terms quickly and accurately; and, has no difficulty solving practical mechanical or electrical problems.

An individual with AVERAGE MECHANICAL COMPREHENSION skills can recognize and understand the relationship of physical forces and mechanical elements in practical situations (e.g., how levers and gears work) with moderate speed and accuracy; is able to learn and figure out mechanical terms with reasonable speed and minimal errors; and, has only a small amount of difficulty solving practical mechanical or electrical problems.

An individual LOW on MECHANICAL COMPREHENSION is unable to recognize and understand the relationship of physical forces and mechanical elements in practical situations; cannot learn and understand mechanical terms; and, has great difficulty solving practical mechanical or electrical problems.

Eye-Limb Coordination

This is the ability to make coordinated movements of any two limbs (e.g., both hands, both feet, or one foot and one hand) with high precision and speed. Eye-limb coordination is most important for tasks where the body is seated or standing while two or more limbs are in motion. For example, eye-limb coordination is very important for flying an airplane, because the pilot must make very precise, coordinated hand and foot movements to maintain the desired altitude and heading. Eye-limb coordination does not apply to tasks in which the movement of other parts of the body must be integrated with arm or leg movements, such as hitting a baseball.

An individual HIGH on EYE-LIMB COORDINATION is very good at doing tasks that require very precise, coordinated movements of two arms, two legs, or an arm and a leg; is very quick to learn and very good at activities like juggling, pinball, sawing a piece of wood in a jigsaw, and operating a lathe; and, has no difficulty doing even very complex tasks.

An individual who has AVERAGE EYE-LIMB COORDINATION is fairly good at activities like juggling, pinball, sawing a piece of wood in a jigsaw, and operating a lathe, but may take some time to initially learn such activities; for very complex tasks, she/he occasionally may have difficulty coordinating arm and leg movements precisely.

An individual LOW on EYE-LIMB COORDINATION may never be able to master tasks requiring coordinated movements of two arms, two legs, or an arm and a leg; even after extended practice, he/she may not be very good at activities like juggling, pinball, sawing a piece of wood in a jigsaw, or operating a lathe.

Precision

This is the ability to make steady, sure hand and arm movements. This ability is important for tracking tasks, such as one might encounter in video games (e.g., pursuing an object that is flying through space or moving through a maze, trying to keep an "X" or a crosshair centered on a moving target). The objects in these tasks may move in either predictable or random patterns. To perform well on these tasks, the individual is required to make precise, smooth hand and arm movements to stay on target. Precision also is important for performing tasks that require the individual to maintain the hand and arm in a single, steady position for a long period of time.

An individual HIGH on PRECISION is able to track targets very accurately, and is capable of holding a hand or arm in a steady position for a long period of time without making even slight tremors or movements.

An individual AVERAGE on PRECISION is able to track targets fairly accurately, but may make occasional large tracking errors; and, can hold a hand or arm in a steady position for a short period of time, but will become significantly less steady after a while.

An individual LOW on PRECISION has a great deal of difficulty tracking targets and makes many tracking errors; and, is not able to hold a hand or arm in a steady position for even a short period of time.

Movement Judgment

This is the ability to judge the relative speed and direction of one or more moving objects. This ability is used to determine where an object will be at a given time, or when two objects will meet or pass each other. For example, movement judgment is important for video games that require shooting or bombing a moving target, because the player must judge when the bomb and target will meet.

An individual HIGH on MOVEMENT JUDGMENT nearly always can determine where a moving object will be at a given time, and nearly always can determine when two moving objects will hit or pass each other, even if the object(s) are moving quickly, at sharp angles, or along a complex curve.

An individual AVERAGE on MOVEMENT JUDGMENT usually can determine where a moving object will be at a given time, but may have some difficulty determining when two moving objects will hit or pass each other if the object(s) are moving very quickly, at sharp angles, or along complex curves.

An individual LOW on MOVEMENT JUDGMENT cannot determine where a moving object will be at a given time, and has great difficulty determining when two moving objects will hit or pass each other, even when the object(s) are moving slowly in relatively straight lines.

Hand and Finger Dexterity

This is the ability to use one's hands or fingers to manipulate very small to moderately large objects with speed and precision.

The individual HIGH on HAND AND FINGER DEXTERITY is able to use the hands and fingers to turn, manipulate, and maintain control of even very small objects, such as screws and delicate jewelry or household ornaments.

The individual with AVERAGE HAND AND FINGER DEXTERITY is able to use the hands and fingers to turn, manipulate, and maintain control of most objects, but may have some difficulty manipulating very small or fine objects.

The individual LOW on HAND AND FINGER DEXTERITY has difficulty using the hands and fingers to manipulate objects--and may be described by others as "klutzy."

Physical Strength

This attribute refers to the ability to apply pressure to and move an object. For example, physical strength is required to lift, push, pull, and carry objects, grip and turn a valve, and perform pull-ups or push-ups.

An individual HIGH on PHYSICAL STRENGTH is able to push and pull even very heavy objects; has no difficulty turning even tight, large valves; can quickly unload heavy bags from a truck; and, can pull herself/himself up a rope to reach a ledge or platform quickly.

An individual with AVERAGE PHYSICAL STRENGTH is able to push and pull moderately heavy objects; may have some difficulty turning large or tight valves; can unload moderately heavy bags from a truck in reasonable time; and, may have some difficulty pulling herself/himself up a rope to reach a ledge or platform.

An individual LOW on PHYSICAL STRENGTH is able to push and pull only light objects; has great difficulty turning valves; cannot unload heavy bags from a truck; and, cannot pull herself/himself up a rope to reach a ledge or platform.

Physical Endurance

This attribute refers to the ability to continue physical activities without getting tired or winded. For example, physical endurance is required for long distance running or swimming, climbing flights of stairs, and shoveling dirt or snow.

An individual HIGH on PHYSICAL ENDURANCE can perform physically active work for long periods of time without getting tired; can perform many repetitions of a physical task without stopping, such as swinging an ax; can climb even long flights of stairs without resting; and, has more stamina than his or her co-workers.

An individual with AVERAGE PHYSICAL ENDURANCE can perform physically active work for moderate periods of time without getting tired; can perform several repetitions of a physical task, such as swinging an ax; can climb a few flights of stairs without resting; and, has about as much stamina as his or her co-workers.

An individual LOW on PHYSICAL ENDURANCE becomes tired after short periods of physically active work; must rest after only a few repetitions of a physical task, such as swinging an ax; must stop to catch his or her breath frequently when climbing stairs; and, has less stamina than his or her co-workers.

Balance and Flexibility

This attribute refers to the ability to keep the body steady while at rest or in motion, and to perform tasks requiring flexing or bending. For example, balance and flexibility are required for walking on a narrow path, adjusting the body to fit through or into a tight area, and stretching or reading to grasp an object.

An individual HIGH on BALANCE AND FLEXIBILITY can keep his/her body steady when standing on a small or slanted surface, such as a ledge or roof; can quickly walk or crawl across a narrow plank or catwalk; can perform tasks while in awkward or uncomfortable positions easily and quickly, such as lying in a tunnel and tightening a bolt overhead; and, never loses his/her balance while bending, twisting, stretching, or reaching.

An individual with AVERAGE BALANCE AND FLEXIBILITY has only minor difficulty keeping his/her body steady when standing on a small or slanted surface, such as a ledge or roof; can generally cross a narrow plank or catwalk without losing balance; may have some difficulty performing tasks while in awkward or uncomfortable positions, such as lying in a tunnel and tightening a bolt overhead; and, generally does not lose his/her balance while bending, twisting, stretching, or reaching.

An individual LOW on BALANCE AND FLEXIBILITY has great difficulty keeping his/her body steady when standing on a small or slanted surface, such as a ledge or roof; cannot cross a narrow plank or catwalk without losing balance; has great difficulty performing tasks while in awkward or uncomfortable positions, such as lying in a tunnel and tightening a bolt overhead; and, usually loses his/her balance while bending, twisting, stretching, or reaching.

Involvement in Athletics

This attribute refers to how often an individual participates in physical activities.

An individual HIGH on INVOLVEMENT IN ATHLETICS participates in individual and team sports and/or exercises vigorously several times per week.

An individual with AVERAGE INVOLVEMENT IN ATHLETICS participates in individual and team sports and/or exercises once or twice per week.

An individual LOW on INVOLVEMENT IN ATHLETICS does not regularly exercise and/or participate in individual and team sports; prefers watching athletic events to being an active participant.

Work Orientation

This is the attribute that reflects a belief that hard work and "sticking to it" pay off. It is characterized by a belief that things that happen to a person are due to one's own efforts (i.e., personal responsibility) rather than chance events, "fate," or what someone else does.

An individual HIGH on WORK ORIENTATION always tries to become excellent at doing his/her job; works very hard; sets extremely high work standards and almost always achieves them; always tries to do a good job even when assigned unpleasant tasks; has remarkably high concentration and really "sticks-to-it" when working on a task; believes firmly that "you get what you deserve" and that people control what happens to them; and, always takes responsibility for her/his work and life.

An individual with AVERAGE WORK ORIENTATION usually tries to become a good worker; works fairly hard; sets moderate work standards and generally tries to do a good job, but may not try as hard when assigned unpleasant tasks; concentrates fairly well and generally "sticks to" a task; believes that people generally control what happens to them but makes allowances for outside influence; and, usually takes responsibility for her/his work and life.

An individual LOW on WORK ORIENTATION does not try to become even an acceptable worker; is not a hard worker; sets less than acceptable work standards; rarely concentrates or "sticks to" a task until completion; does not believe that "you get what you deserve," rather believes that what happens to people is beyond their personal control; and, generally does not take responsibility for her/his work and life.

Sociability

This attribute reflects how well one gets along in social situations. It includes being outgoing, talkative, and friendly, and being comfortable in group situations.

The individual HIGH on SOCIABILITY is outgoing, talkative, relates easily to others, and is very responsive and expressive in social environments; has many friends and makes new friends easily; and, is comfortable in group situations.

The individual AVERAGE on SOCIABILITY is somewhat outgoing and talkative, but is occasionally quiet and a little shy in large groups; may find it difficult to think of things to talk about around new friends and acquaintances; and, neither seeks out or avoids group situations.

The individual LOW on SOCIABILITY is quiet, reserved, and shy; generally avoids getting together with large groups; typically has only a few close friends; and, typically prefers spending time alone.

Cooperation/Stability

This is the tendency to be pleasant in dealing with others, to be tolerant, tactful, and helpful, and to maintain control over one's emotions, even when provoked.

The individual HIGH on COOPERATION/STABILITY is very pleasant to deal with and is calm even during stressful situations; is tactful, tolerant, helpful, and very easy to get along with; does not become upset or hostile when criticized or teased; and, maintains an even, generally happy mood almost all of the time.

The individual AVERAGE on COOPERATION/STABILITY generally is pleasant to deal with, but may become unpleasant and tense during very stressful situations; is tactful and easy to get along with most of the time, but may occasionally become upset or hostile when criticized or teased; and, usually maintains an even mood.

The individual LOW on COOPERATION/STABILITY is generally very difficult to deal with, especially during stressful situations; lacks tact, is intolerant of faults in others, is not very helpful to friends or acquaintances, and does not get along well with others; becomes extremely upset or hostile when criticized; has frequent changes in mood; and, often becomes very angry and upset for no apparent reason.

Energy

This attribute refers to the level of energy and enthusiasm shown by a person.

An individual HIGH on ENERGY has the energy and vitality to get things done; is nearly always enthusiastic, optimistic, and cheerful; and, has an active life-style.

An individual with AVERAGE ENERGY has the energy and vitality to get most of her/his responsibilities completed, but may not have the energy to work extra hours or meet an emergency; occasionally lacks enthusiasm; is sometimes pessimistic and depressed; and, has a somewhat active life-style.

An individual LOW on ENERGY often does not have the energy to complete assigned tasks; is generally pessimistic and depressed; is rarely enthusiastic; and, leads an inactive life.

Conscientiousness

This is the attribute that reflects respect for discipline, order, structure, regulations, and authority. It results in planful, dependable, well-organized behavior, stability, and resistance to change.

An individual HIGH on CONSCIENTIOUSNESS accepts authority and the need for and value of discipline; is conventional regarding social and political views; resists social change; seeks order in her/his life; has a high degree of self-control; thinks before acting; and, is always honest and law-abiding.

An individual AVERAGE on CONSCIENTIOUSNESS generally accepts authority and the value of discipline; is not extreme in social and political views; is willing to work with the system to affect planned social change; typically plans prior to taking action but can be spontaneous; has moderate self-control; and, is generally honest and law-abiding.

An individual LOW on CONSCIENTIOUSNESS does not accept authority or the need for and value of discipline; is unconventional in his/her social and political views; enjoys change and dislikes order in her/his life; lacks self-control; acts before thinking; may have a history of problems with schools and law enforcement agencies; and, tends to be dishonest.

Dominance/Confidence

This is the tendency to seek positions of leadership and influence over others. It reflects confidence in one's abilities and an expectation that one can succeed.

The individual HIGH on DOMINANCE/CONFIDENCE enjoys being a leader and seeks leadership positions; can be highly forceful and persuasive when the situation requires; has succeeded in many previous undertakings and is very confident of success in the future.

The individual with AVERAGE DOMINANCE/CONFIDENCE may enjoy being a leader, but generally is not persuasive; has moderate confidence in his/her abilities; and, is usually confident of success in the future.

The individual LOW on DOMINANCE/CONFIDENCE does not seek leadership positions; is timid about offering opinions, advice, or suggestions; often feels incapable; and, has many doubts about succeeding in the future.

Interest in Using Tools and Machines

This attribute reflects a preference for physically active work and practical, concrete activities. Individuals who express this attribute enjoy maintaining and repairing machines or engines, driving cars or trucks, and constructing buildings or roads.

Interest in Rugged Activities

This attribute reflects a preference for rugged activities, working outdoors, and using a weapon. Individuals who express this attribute enjoy hunting, fishing, camping, and hiking in the wilderness.

Interest in Protective Services

This attribute reflects a preference for activities that involve protecting people and property from harm. Individuals who express this attribute enjoy guarding property, fighting fires, enforcing laws, and promoting safety.

Interest in Technical Activities

This attribute reflects a preference for activities that involve technical equipment or skills. Individuals who express this attribute enjoy reading and drawing blueprints and reading about electronics. These individuals like to build, repair, and operate radios, and would enjoy operating movie or television cameras and sound equipment.

Interest in Science

This attribute reflects a preference for scholarly, intellectual, and scientific activities. Individuals who express this attribute enjoy reading scientific or scholarly material and prefer to work independently. These individuals enjoy doing research, and like to think about and solve puzzling problems, such as diagnosing a disease or finding an error in a computer program.

Interest in Leadership

This attribute reflects a preference for teaching, helping, persuading, and leading others. Individuals who express this attribute like to be "in charge" and to organize a group and mold it into an efficient team. These individuals enjoy giving speeches, inspiring others to excel, and winning others over to their viewpoint, and are likely to be good listeners.

Interest in Artistic Activities

This attribute reflects a preference for unstructured, expressive, creative activities. Individuals who express this attribute enjoy reading poetry and short stories. These individuals prefer to listen to classical music, enjoy ballet, and like to visit art museums.

Interest in Efficiency and Organization

This attribute reflects a preference for well-ordered and systematic activities. Individuals who express this attribute enjoy keeping records, ordering and keeping track of supplies, and working with tables of numbers. These individuals like to follow a list of steps in completing a task. They enjoy preparing and serving food, and keeping things clean and orderly.

PERFORMANCE DEFINITION

In completing your ratings, five general areas of performance should be kept in mind: (1) core technical proficiency, (2) general soldiering proficiency, (3) effort and leadership, (4) personal discipline, and (5) physical fitness and military bearing. These performance areas are defined below.

Core Technical Proficiency: This peformance area represents the proficiency with which the soldier performs the tasks that are "central" to the MOS. The tasks represent the core of the job and are the primary definers of the MOS. This performance area does not include the individual's willingness to perform the task or the degree to which the individual can coordinate efforts with others. It refers to how well the individual can execute the core technical tasks the job requires, given a willingness to do so.

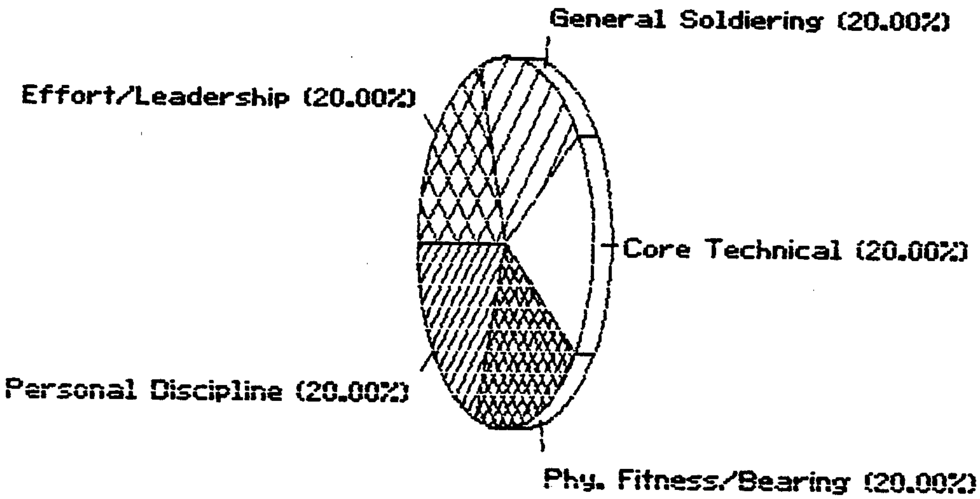
General Soldiering Proficiency: In addition to the core technical content specific to an MOS, individuals in every MOS also are responsible for being able to perform a variety of general soldiering tasks. For example, determines grid coordinates on military maps; puts on, wears, and removes M17 series protective mask with hood; determines a magnetic azimuth using a compass; collects and reports information; and recognizes and identifies friendly and threat aircraft. Performance of this area represents overall proficiency on these and other general soldiering tasks. Again, it refers to how well the individual can execute general soldiering tasks, given a willingness to do so.

Effort and Leadership: This performance area reflects the degree to which the individual exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers. That is, can the individual be counted on to carry out assigned tasks, even under adverse conditions, to exercise good judgment, and to be generally dependable and proficient? While appropriate knowledges and skills are necessary for successful performance, this area is only meant to reflect the individual's willingness to do the job required and to be cooperative and supportive with other soldiers.

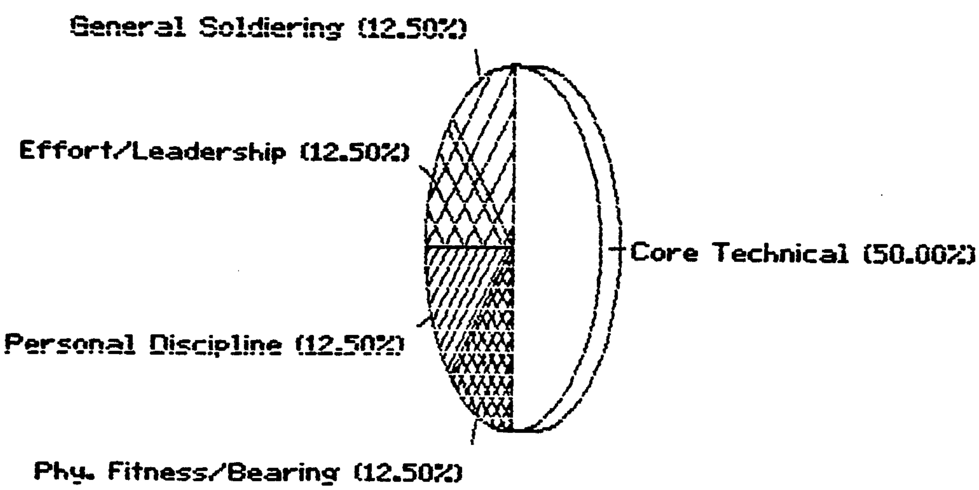
Personal Discipline: This performance area reflects the degree to which the individual adheres to Army regulations and traditions, exercises personal self-control, demonstrates integrity in day-to-day behavior, and does not create disciplinary problems. People who rank high on this area show a commitment to high standards of personal conduct.

Physical Fitness and Military Bearing: This performance area represents the degree to which the individual maintains an appropriate military appearance and bearing and stays in good physical condition.

Areas of Job Performance for Soldiers



Areas of Job Performance for Soldiers



11B10 - INFANTRYMAN

Battlefield Survival

Performs preventive maintenance and assists in organizational maintenance on weapons and equipment.
Protects self, weapons, and equipment from chemical and other contaminants.
Carries and prepares ammunition for use and loads weapons.
Administers first aid.
Applies field sanitation methods.
Insures highest state of physical readiness at all times.

Combat Techniques and Tactics

Lays field mines.
Performs basic communications funtions and operates platoon communications equipment.
Employs cover, concealment, and camouflage.
Assists in construction of fortifications and barriers, including mine fields.
Prepares simple demolitions.
Assists in breaching and clearing mine fields and obstacles.
Performs indiredt and aerial fire support.
Captures prisoners and renders verbal reports on activities.
Reacts to oral commands and signals.
Applies principles of escape and evasion.
Collects and reports tactical information as a member of combat or reconnaissance patrol.
Performs tactical associated duties in support of offensive, defensive, and specialized missions.

Weapons

Uses individual infantry weapons and associated equipment.
Carries, emplaces, sights, and fires machine gun.
Throws grenades and operates grenade launcher.
Engages in hand-to-had combat and employs bayonet and silent weapons.
Carriers, sights, and fires light antiarmor weapons.

Tactical Vehicle/Support Activities

Operates wheeled and tracked truck to transport personnel, supplies, and equipment.
Performs and assists in unit and organizational maintenance.
Performs duties as guard.
Delivers messages and performs other elementary tasks in support of operations and intelligence functions.
Performs drill and ceremonies and other post, camp, and station duties as required.

Maintenance

Conducts preventive maintenance checks and services (PMCS) on organizational equipment.

Driving

Operates track vehicles over varied terrain in varied visibility and weather conditions, using techniques of movement commensurate with the threat.
Reacts to audio and visual signals.
Assists in refueling and vehicle operations.

63B10 - LIGHT-WHEEL VEHICLE MECHANIC

Repairs and Maintenance

Performs repairs and preventive maintenance checks and services in accordance with appropriate maintenance allocation chart (MAC).
Tests lead acid batteries for serviceability.
Prepares vehicles and equipment for operation under abnormal conditions by sealing, waterproofing, and servicing with special fuels, coolants, and lubricants.
Employs unit maintenance techniques to correct malfunctions of mechanical, electrical, and hydraulic brake systems or components.
Performs preventive maintenance on tools, repair parts, and equipment.
Adjusts operating mechanisms.
Tunes engines and replaces repair parts and components as outlined in the appropriate MAC.
Tests and operates vehicles and equipment as required and observes for evidence of abnormal operation.
Operates, or assists in the operation of wheel vehicles in conjunction with maintenance and recovery operations.
Employs applicable TMDE in conjunction with technical publications to troubleshoot and test continuity of electrical circuits.
Interprets schematic diagrams.
Determines extent of corrective action and repair parts required to correct malfunctions.
Uses hand tools.
Operates power tools and equipment.
Obtains replacement parts from maintenance parts supply section.
Turns in unserviceable parts to maintenance parts supply section while adhering to authorized policies and required procedures.
Reads, understands, and applies information contained in unit maintenance technical manuals and common schematic diagrams.
Knows and applies required safety procedures.
Performs or assists in vehicle recovery and evacuation of vehicles that weigh 5 tons or less.
Maneuvers wrecker to desired position.
Uses recovery rigging techniques to attach and secure hoists, cables, and slings, using authorized equipment and effects recovery operations.
Performs basic oxyacetylene cutting.
Performs operator maintenance on light-wheel wreckers.
Knows and applies recovery operations safety procedures.
Reads topographic maps.
Uses proper tactical radio procedures.

71L10 - ADMINISTRATIVE SPECIALIST

Typing

Types military and nonmilitary correspondence, messages, orders, recurring, and special reports, requisitions, forms, regulations, directives, SOP, and similar materials.
Types material in draft and final copy.
Uses written notes, drafts, verbal instructions, or other sources to prepare copy.
Prepares typewritten copy in proper format to comply with prescribed correspondence directives, including the use of authorized abbreviations.
Employs basic principles of English composition and grammar in preparing correspondence.
Proofreads typed material against source documents.
Assembles final product review, signature, authentication, or other disposition.
Maintains typewriters.

Clerical

Greets office visitors and directs visitors to proper destinations.
Announces arrival of visitors.
Answers telephone and records telephone messages.
Opens, sorts, routes, and delivers incoming correspondence and messages.
Prepares suspense control documents and maintains suspense files.
Reproduces copies of printed material using copying equipment.
Signs receipts for and picks up registered and certified mail.
Prepares file labels and guides.
Determines proper functional file and file numbers, posts to documents, and files documents in accordance with TAFFS.
Destroys or disposes of files in accordance with disposition instructions.
Receives publications and maintains publications library.
Posts changes to regulations and directives.
Receives and stocks blank forms.

INSTRUCTIONS FOR VALIDITY JUDGMENT

SECTION I: INTRODUCTION TO THE JUDGMENT TASK

Today you will be asked to make judgments about attributes, job performance, and the relationship between the two. Let's start with some key definitions.

Attribute: A knowledge, skill, ability, or other personal characteristic such as interests, values, temperament, etc. Persons can be tested or evaluated and assigned scores on attributes (e.g., reading comprehension, numerical ability).

Job Performance: The way that a person carries out the job tasks and other requirements of the job. Persons can be tested or evaluated and assigned scores on various areas of job performance (e.g., physical fitness test scores, AIT grades, instructors' and supervisors' evaluations).

Validity: The way in which scores on an attribute are related to scores on job performance. For example, if persons scoring low on an attribute also score low on job performance, and persons scoring high on that same attribute also score high on job performance, then that attribute has high validity. But if scores on an attribute are not related to scores on job performance, then there is little or no validity for that attribute. For example, if we find that almost all persons who score in the top twenty percent on a writing skills test are also consistently the best newspaper reporters, as rated by their supervisors, and low scorers on the test are consistently rated the worst reporters, then writing skill is valid for newspaper reporting.

You have been selected to make these judgments because you are familiar with Army job performance. In completing your judgments, think about the performance of soldiers you have supervised in the last year or so. You will be given a separate handout defining job performance in more detail. Please refer to it when making your judgments.

You will be asked to make judgments about several attributes. These attributes are defined in a separate handout, and descriptions of individuals scoring high, average, and low on each attribute are provided. The attributes include cognitive ability, temperament and background, interests, and values.

SECTION II: SPECIFIC INSTRUCTIONS

For each attribute, judge the validity of the attribute for performance in each of the five job areas outlined in the "Performance Definition" handout.

As discussed above, validity indicates the relationship between scores on a test of the attribute and scores on job performance. For example, if scores on the attribute are not related to job performance scores, validity would be very low. This is the case when individuals who all have high scores on the test have job performance scores ranging from very low to very high. In this low validity situation, we can tell nothing about how a soldier might do on the job even though we do know his or her score on the attribute test. If scores on the attribute are closely related to job performance scores, validity would be high. This is the case when scores on the test very closely match scores on job performance for nearly all individuals. That is, nearly all those getting high scores on the attribute test also get high scores on job performance and nearly all those getting low scores on the test get low scores on job performance. In this high validity situation, we can tell a whole lot about how well a soldier will do on the job when we know what score she/he received on the test of the attribute.

Use the following scale to make the validity judgments (this scale is also shown on a separate handout):

VALIDITY JUDGMENT SCALE

Response Option	Level of Validity
0	no validity
1	
2	low validity
3	
4	medium validity
5	
6	high validity
7	
8	extremely high validity

Please refer to the sample response sheet below as you read through the following examples for Verbal Ability and Reasoning.

	Validity for Job Area				
Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Verbal Ability	6	4	2	0	1
Reasoning	4	3	2	3	1

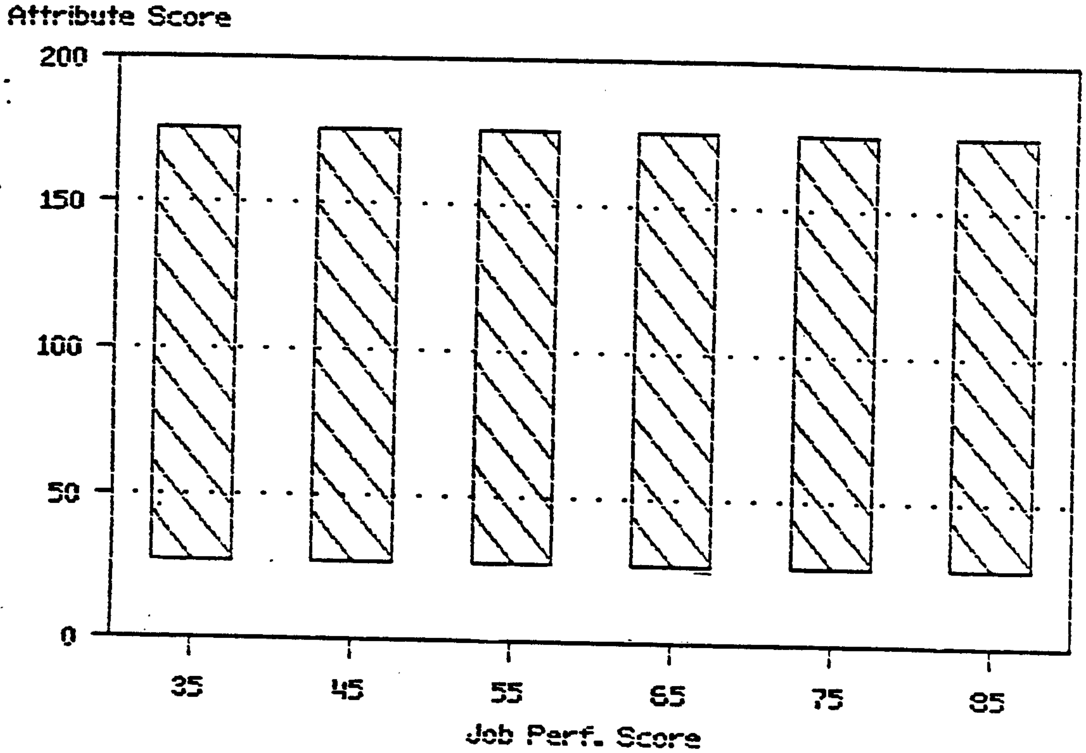
Joe supervises 95B (Military Police). After thoroughly reading the performance definition and the definition of Verbal Ability, Joe believes that scores on a Verbal Ability test are probably highly related to performance of the core technical tasks of a 95B. So, he enters a 6 in the column headed "Validity for Job Area: Core Technical." Joe believes that scores on a Verbal Ability test are probably moderately related to performance scores on general soldiering tasks, so he enters a 4 in the column headed "Validity for Job Area: General Soldiering." Joe completes the rest of the job area validity judgments for Verbal Ability in a similar manner.

Joe then proceeds to the second attribute, Reasoning. After reading the definition of Reasoning, Joe believes that scores on a Reasoning test are probably moderately related to core technical task performance for 95Bs. So, he enters a 4 in the column headed "Validity for Job Area: Core Technical." Joe thinks scores on a Reasoning test are probably somewhat related to general soldiering performance scores, so he enters a 3 in the column headed "Validity for Job Area: General Soldiering." Joe completes the rest of the job area validity judgments in a similar manner.

Keep the "Performance Definition" and the rating scale handy and refer to them as frequently as necessary while making your judgments.

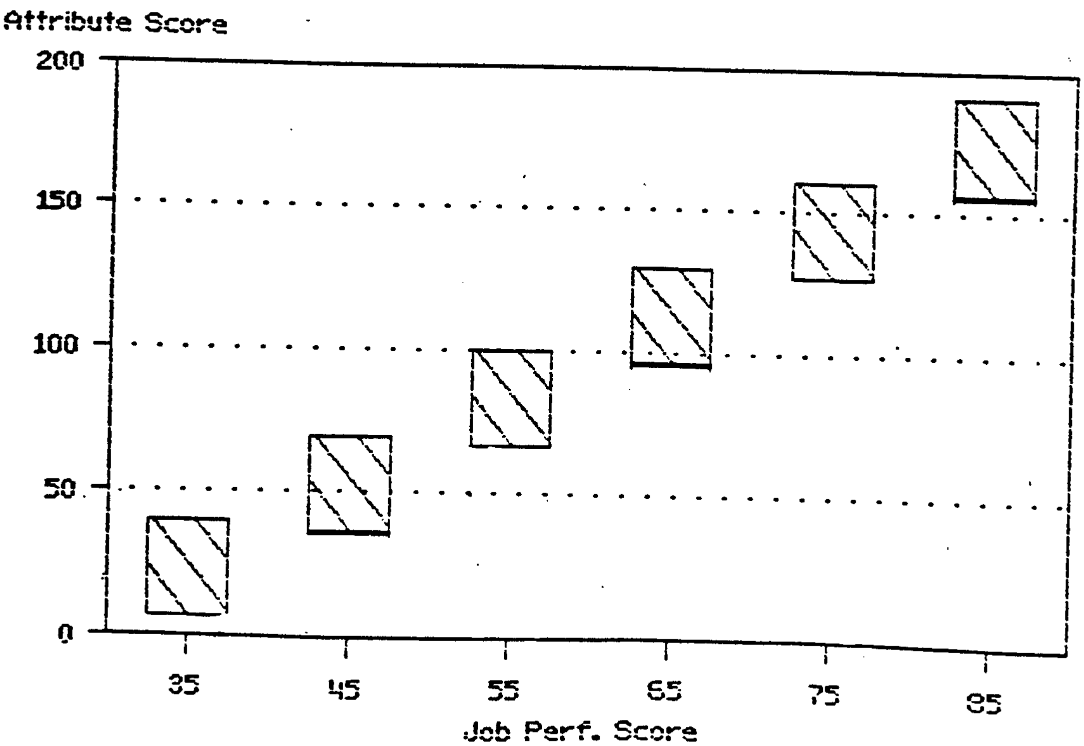
Graph of Low Validity Attribute

Relationship of Attribute to Performance



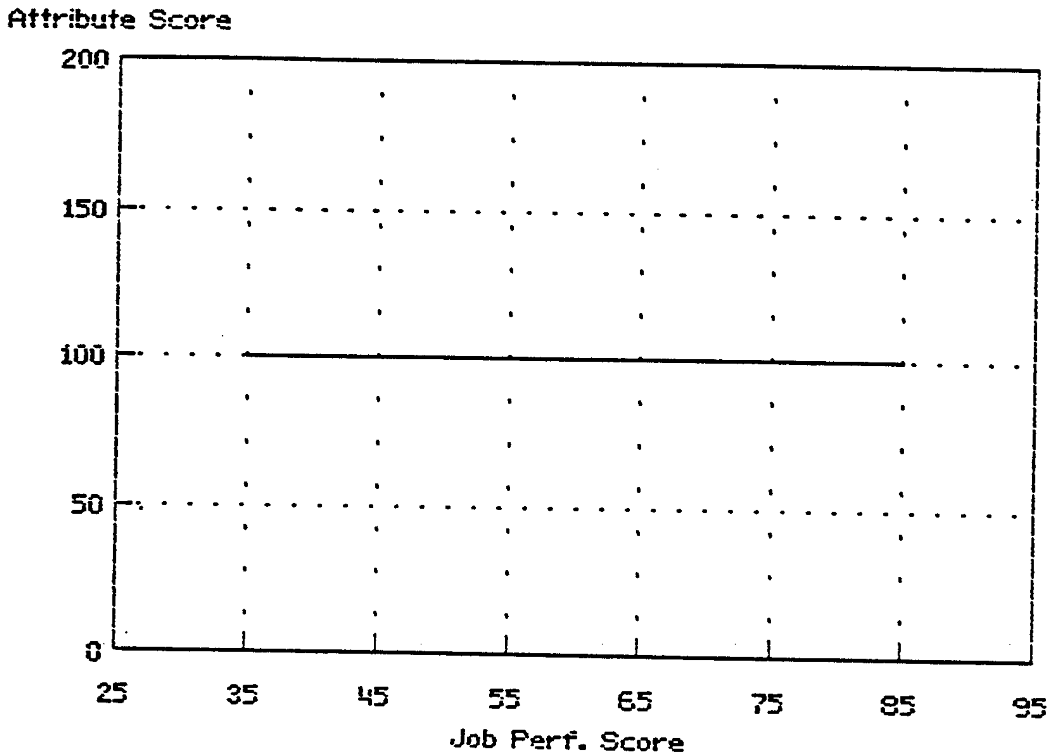
Graph of High Validity Attribute

Relationship of Attribute to Performance



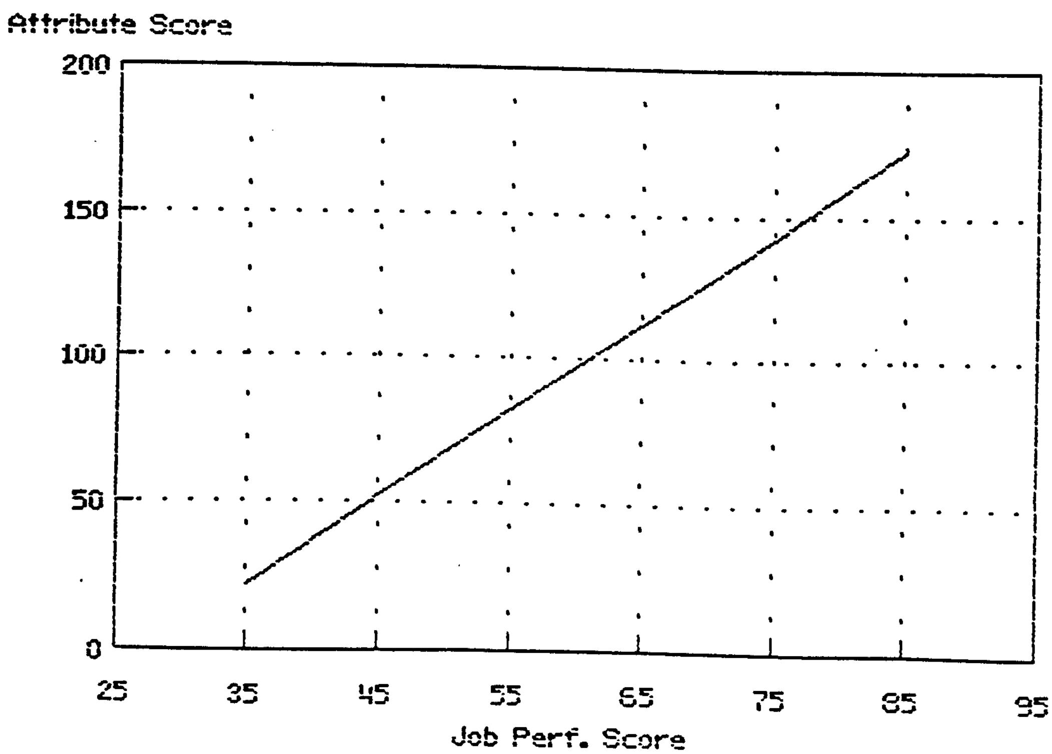
Graph of Low Validity Attribute

Relationship of Attribute to Performance



Graph of High Validity Attribute

Relationship of Attribute to Performance



Rated MOS ________ Your Rank ________ Today's Date ________ Name ______________________

RESPONSE SHEET FOR VALIDITY JUDGMENT: COGNITIVE ABILITIES

Attribute	Validity for Job Area: Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Verbal Ability	___	___	___	___	___
Reasoning	___	___	___	___	___
Number Ability	___	___	___	___	___
Spatial Ability	___	___	___	___	___
Closure	___	___	___	___	___
Mental Information Processing	___	___	___	___	___
Perceptual Speed and Accuracy	___	___	___	___	___
Memory	___	___	___	___	___
Mechanical Comprehension	___	___	___	___	___

RESPONSE SHEET FOR VALIDITY JUDGMENT: COORDINATION AND PHYSICAL ABILITIES

Validity for Job Area

Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Eye-Limb Coordination	___	___	___	___	___
Precision	___	___	___	___	___
Movement Judgment	___	___	___	___	___
Hand and Finger Dexterity	___	___	___	___	___
Physical Strength	___	___	___	___	___
Physical Endurance	___	___	___	___	___
Balance and Flexibility	___	___	___	___	___
Involvement in Athletics	___	___	___	___	___

RESPONSE SHEET FOR VALIDITY JUDGMENT: TEMPERAMENT/ATTITUDE

Validity for Job Area

Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Work Orientation	___	___	___	___	___
Sociability	___	___	___	___	___
Cooperation/Stability	___	___	___	___	___
Energy	___	___	___	___	___
Conscientiousness	___	___	___	___	___
Dominance/Confidence	___	___	___	___	___

RESPONSE SHEET FOR VALIDITY JUDGMENT: INTERESTS

Attribute	Validity for Job Area: Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Interest in Using Tools and Machines	___	___	___	___	___
Interest in Rugged Activities	___	___	___	___	___
Interest in Protective Services	___	___	___	___	___
Interest in Technical Activities	___	___	___	___	___
Interest in Science	___	___	___	___	___
Interest in Leadership	___	___	___	___	___
Interest in Artistic Activities	___	___	___	___	___
Interest in Efficiency and Organization	___	___	___	___	___

Rated MOS _______ Your Rank _______ Today's Date _______ Name ____________________

RESPONSE SHEET FOR VALIDITY JUDGMENT: COGNITIVE ABILITIES

Attribute	Validity for Job Component									
Verbal Ability	___	___	___	___	___	___	___	___	___	___
Reasoning	___	___	___	___	___	___	___	___	___	___
Number Ability	___	___	___	___	___	___	___	___	___	___
Spatial Ability	___	___	___	___	___	___	___	___	___	___
Closure	___	___	___	___	___	___	___	___	___	___
Mental Information Processing	___	___	___	___	___	___	___	___	___	___
Perceptual Speed and Accuracy	___	___	___	___	___	___	___	___	___	___
Memory	___	___	___	___	___	___	___	___	___	___
Mechanical Comprehension	___	___	___	___	___	___	___	___	___	___

RESPONSE SHEET FOR VALIDITY JUDGMENT: COORDINATION AND PHYSICAL ABILITIES

Validity for Job Component

Attribute										
Eye-Limb Coordination	___	___	___	___	___	___	___	___	___	___
Precision	___	___	___	___	___	___	___	___	___	___
Movement Judgment	___	___	___	___	___	___	___	___	___	___
Hand and Finger Dexterity	___	___	___	___	___	___	___	___	___	___
Physical Strength	___	___	___	___	___	___	___	___	___	___
Physical Endurance	___	___	___	___	___	___	___	___	___	___
Balance and Flexibility	___	___	___	___	___	___	___	___	___	___
Involvement in Athletics	___	___	___	___	___	___	___	___	___	___

RESPONSE SHEET FOR VALIDITY JUDGMENT: TEMPERAMENT/ATTITUDE

Validity for Job Component

Attribute										
Work Orientation										
Sociability										
Cooperation/Stability										
Energy										
Conscientiousness										
Dominance/Confidence										

RESPONSE SHEET FOR VALIDITY JUDGMENT: INTERESTS

Attribute	Validity for Job Component									
Interest in Using Tools and Machines	___	___	___	___	___	___	___	___	___	___
Interest in Rugged Activities	___	___	___	___	___	___	___	___	___	___
Interest in Protective Services	___	___	___	___	___	___	___	___	___	___
Interest in Technical Activities	___	___	___	___	___	___	___	___	___	___
Interest in Science	___	___	___	___	___	___	___	___	___	___
Interest in Leadership	___	___	___	___	___	___	___	___	___	___
Interest in Artistic Activities	___	___	___	___	___	___	___	___	___	___
Interest in Efficiency and Organization	___	___	___	___	___	___	___	___	___	___

INSTRUCTIONS FOR ATTRIBUTE RANKING: VALIDITY

In this task, you will be asked to rank order the thirty-one attributes according to their validity for *overall* performance in the MOS you supervise. In considering overall performance, think about all areas of job performance. You may want to review the performance definition and the attribute definitions before beginning.

Look at the list of thirty-one attributes. Think about how valid each attribute is for overall performance in the MOS you supervise. Pick the attribute that you think is most valid and place a "1" in the space beside the attribute. Next, pick the attribute that you think is least valid and place a "31" beside the attribute. Then, pick the attribute that you think is second most valid and place a "2" beside the attribute. Next, pick the attribute that you think is the second least valid and place a "30" to the left of the attribute. Complete the rest of the rank orderings in this way, alternating back and forth from most to least valid, until you have ranked all thirty-one attributes. When you are finished, review the rank orders to be certain that you have used each number (1 through 31) only once.

Please refer to the diagrams as you go through the following example.

A. Joe supervises 95B, Military Police. After reviewing the attribute definitions and the definition of job performance, Joe decides that Physical Strength is the most valid attribute for overall performance as a 95B. So, Joe puts a "1" beside Physical Strength.

___ Verbal Ability
___ Reasoning
___ Number Ability
___ Spatial Ability
___ Closure
___ Mental Information Processing
___ Perceptual Speed and Accuracy
___ Memory
___ Mechanical Comprehension
___ Eye-Limb Coordination
___ Precision

___ Movement Judgment
___ Hand and Finger Dexterity
_1_ Physical Strength
___ Physical Endurance
___ Balance and Flexibility
___ Involvement in Athletics
___ Work Orientation
___ Sociability
___ Cooperation/Stability
___ Energy
___ Conscientiousness
___ Dominance/Confidence

___ Interest in Using Tools and Machines
___ Interest in Rugged Activities
___ Interest in Protective Services
___ Interest in Technical Activities
___ Interest in Science
___ Interest in Leadership
___ Interest in Artistic Activities
___ Interest in Efficiency and Organization

B. Next, Joe decides that Interest in Artistic Activities is <u>least</u> valid for overall performance. So, he puts a "31" beside Interest in Artistic Activities.

___ Verbal Ability
___ Reasoning
___ Number Ability
___ Spatial Ability
___ Closure
___ Mental Information Processing
___ Perceptual Speed and Accuracy
___ Memory
___ Mechanical Comprehension
___ Eye-Limb Coordination
___ Precision

___ Movement Judgment
___ Hand and Finger Dexterity
1 Physical Strength
___ Physical Endurance
___ Balance and Flexibility
___ Involvement in Athletics
___ Work Orientation
___ Sociability
___ Cooperation/Stability
___ Energy
___ Conscientiousness
___ Dominance/Confidence

___ Interest in Using Tools and Machines
___ Interest in Rugged Activities
___ Interest in Protective Services
___ Interest in Technical Activities
___ Interest in Science
___ Interest in Leadership
31 Interest in Artistic Activities
___ Interest in Efficiency and Organization

C. Joe then decides that Verbal Ability is the <u>second most</u> valid attribute for overall performance. He puts a "2" beside Verbal Ability.

2 Verbal Ability
___ Reasoning
___ Number Ability
___ Spatial Ability
___ Closure
___ Mental Information Processing
___ Perceptual Speed and Accuracy
___ Memory
___ Mechanical Comprehension
___ Eye-Limb Coordination
___ Precision

___ Movement Judgment
___ Hand and Finger Dexterity
1 Physical Strength
___ Physical Endurance
___ Balance and Flexibility
___ Involvement in Athletics
___ Work Orientation
___ Sociability
___ Cooperation/Stability
___ Energy
___ Conscientiousness
___ Dominance/Confidence

___ Interest in Using Tools and Machines
___ Interest in Rugged Activities
___ Interest in Protective Services
___ Interest in Technical Activities
___ Interest in Science
___ Interest in Leadership
31 Interest in Artistic Activities
___ Interest in Efficiency and Organization

D. Next, Joe decides that Closure is the <u>second least</u> valid attribute, so he puts a "30" beside Closure.

2 Verbal Ability

___ Reasoning

___ Number Ability

___ Spatial Ability

30 Closure

___ Mental Information Processing

___ Perceptual Speed and Accuracy

___ Memory

___ Mechanical Comprehension

___ Eye-Limb Coordination

___ Precision

___ Movement Judgment

___ Hand and Finger Dexterity

1 Physical Strength

___ Physical Endurance

___ Balance and Flexibility

___ Involvement in Athletics

___ Work Orientation

___ Sociability

___ Cooperation/Stability

___ Energy

___ Conscientiousness

___ Dominance/Confidence

___ Interest in Using Tools and Machines

___ Interest in Rugged Activities

___ Interest in Protective Services

___ Interest in Technical Activities

___ Interest in Science

___ Interest in Leadership

31 Interest in Artistic Activities

___ Interest in Efficiency and Organization

Joe continues to rank order the attributes in this way, alternating back and forth between most and least valid, until all thirty-one attributes have been rank-ordered. As a final check, Joe makes certain that he used each number from 1 through 31 only once.

ATTRIBUTE RANKING WORKSHEET

___ Verbal Ability

___ Reasoning

___ Number Ability

___ Spatial Ability

___ Closure

___ Mental Information Processing

___ Perceptual Speed and Accuracy

___ Memory

___ Mechanical Comprehension

___ Eye-Limb Coordination

___ Precision

___ Movement Judgment

___ Hand and Finger Dexterity

___ Physical Strength

___ Physical Endurance

___ Balance and Flexibility

___ Involvement in Athletics

___ Work Orientation

___ Sociability

___ Cooperation/Stability

___ Energy

___ Conscientiousness

___ Dominance/Confidence

___ Interest in Using Tools and Machines

___ Interest in Rugged Activities

___ Interest in Protective Services

___ Interest in Technical Activities

___ Interest in Science

___ Interest in Leadership

___ Interest in Artistic Activities

___ Interest in Efficiency and Organization

Please check to see that you used each number from 1 through 31 only once.

INSTRUCTIONS FOR IMPORTANCE JUDGMENT

SECTION I: INTRODUCTION TO THE JUDGMENT TASK

Today you will be asked to make judgments about attributes, job performance, and the relationship between the two. Let's start with some key definitions.

Attribute: A knowledge, skill, ability, or other personal characteristic such as interests, values, temperament, etc. Persons can be tested or evaluated and assigned scores on attributes (e.g., reading comprehension, numerical ability).

Job Performance: The way that a person carries out the job tasks and other requirements of the job. Persons can be tested or evaluated and assigned scores on various areas of job performance (e.g., physical fitness test scores, AIT grades, instructors' and supervisors' evaluations).

Importance: The degree to which use of an attribute is directly involved in job performance. For example, auto mechanics must know and make use of information about how internal combustion engines work.

You have been selected to make these judgments because you are familiar with Army job performance. In completing your judgments, think about the performance of soldiers you have supervised in the last year or so. You will be given a separate handout defining job performance in more detail. Please refer to it when making your judgments.

You will be asked to make judgments about several attributes. These attributes are defined in a separate handout, and descriptions of individuals scoring high, average, and low on each attribute are provided. The attributes include cognitive ability, temperament and background, interests, and values.

SECTION II: SPECIFIC INSTRUCTIONS

For each attribute, rate the importance of the attribute for performance in each of the five job areas outlined in the "Performance Definition" handout. Use the following scale for the importance judgments (this scale is also shown on a separate handout):

IMPORTANCE JUDGMENT SCALE

Response Option	Level of Importance
0	no importance
1	
2	low importance
3	
4	moderate importance
5	
6	high importance
7	
8	extremely high importance

Please refer to the sample response sheet below as you read through examples for Verbal Ability and Reasoning.

	Importance for Job Area				
Attribute	**Core Technical**	**General Soldiering**	**Effort and Leadership**	**Personal Discipline**	**Fitness and Bearing**
Verbal Ability	6	4	2	0	1
Reasoning	4	3	2	3	1

Joe supervises 95B (Military Police). After thoroughly reading the performance definition and the definition of Verbal Ability, Joe concludes that Verbal Ability is of high importance for success on the core technical tasks a 95B must perform. So, Joe enters response option 6, for "High Importance," under the column headed "Importance to Job Area: Core Technical." Joe thinks Verbal Ability is moderately important for general soldiering tasks, so he enters option 4, signifying "Moderate Importance," under the column headed "Importance to Job Area: General Soldiering." Joe thinks that Verbal Ability is only slightly important to effort and leadership, so he enters a 2 in the column headed "Importance to Job Area: Effort & Leadership." Joe completes the rest of the judgments for the five job areas in a similar manner.

Joe then proceeds to the second attribute, Reasoning. After reading the definition of Reasoning, Joe decides that Reasoning is of moderate importance for performance of the 95B core technical tasks. So, he enters a 4 in the column headed "Importance to Job Area: Core Technical." Joe thinks Reasoning is somewhat important for general soldiering tasks, so he enters a 3 in the column headed "Importance to Job Area: General Soldiering." Joe believes Reasoning is of low importance for effort and leadership, so he enters a 2 in the column headed "Importance to Job Area: Effort & Leadership." Joe completes the rest of the importance judgments for the five job areas in a similar manner.

Keep the "Performance Definition" and rating scale handy and refer to them as frequently as necessary while making your judgments.

Rated MOS ______ Your Rank ______ Today's Date ______ Name ______________

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: COGNITIVE ABILITIES

	Importance for Job Area				
Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Verbal Ability	___	___	___	___	___
Reasoning	___	___	___	___	___
Number Ability	___	___	___	___	___
Spatial Ability	___	___	___	___	___
Closure	___	___	___	___	___
Mental Information Processing	___	___	___	___	___
Perceptual Speed and Accuracy	___	___	___	___	___
Memory	___	___	___	___	___
Mechanical Comprehension	___	___	___	___	___

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: COORDINATION AND PHYSICAL ABILITIES

Importance for Job Area

Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Eye-Limb Coordination	___	___	___	___	___
Precision	___	___	___	___	___
Movement Judgment	___	___	___	___	___
Hand and Finger Dexterity	___	___	___	___	___
Physical Strength	___	___	___	___	___
Physical Endurance	___	___	___	___	___
Balance and Flexibility	___	___	___	___	___
Involvement in Athletics	___	___	___	___	___

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: TEMPERAMENT/ATTITUDE

Importance for Job Area

Attribute	Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Work Orientation					
Sociability					
Cooperation/Stability					
Energy					
Conscientiousness					
Dominance/Confidence					

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: INTERESTS

Attribute	Importance for Job Area: Core Technical	General Soldiering	Effort and Leadership	Personal Discipline	Fitness and Bearing
Interest in Using Tools and Machines	___	___	___	___	___
Interest in Rugged Activities	___	___	___	___	___
Interest in Protective Services	___	___	___	___	___
Interest in Technical Activities	___	___	___	___	___
Interest in Science	___	___	___	___	___
Interest in Leadership	___	___	___	___	___
Interest in Artistic Activities	___	___	___	___	___
Interest in Efficiency and Organization	___	___	___	___	___

Rated MOS ________ Your Rank ________ Today's Date ________ Name ________________________________

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: COGNITIVE ABILITIES

Importance for Job Component

Attribute										
Verbal Ability										
Reasoning										
Number Ability										
Spatial Ability										
Closure										
Mental Information Processing										
Perceptual Speed and Accuracy										
Memory										
Mechanical Comprehension										

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: COORDINATION AND PHYSICAL ABILITIES

Attribute	Importance for Job Component									
Eye-Limb Coordination	___	___	___	___	___	___	___	___	___	___
Precision	___	___	___	___	___	___	___	___	___	___
Movement Judgment	___	___	___	___	___	___	___	___	___	___
Hand and Finger Dexterity	___	___	___	___	___	___	___	___	___	___
Physical Strength	___	___	___	___	___	___	___	___	___	___
Physical Endurance	___	___	___	___	___	___	___	___	___	___
Balance and Flexibility	___	___	___	___	___	___	___	___	___	___
Involvement in Athletics	___	___	___	___	___	___	___	___	___	___

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: TEMPERAMENT/ATTITUDE

Importance for Job Component

Attribute										
Work Orientation										
Sociability										
Cooperation/Stability										
Energy										
Conscientiousness										
Dominance/Confidence										

RESPONSE SHEET FOR IMPORTANCE JUDGMENT: INTERESTS

Attribute	Importance for Job Component									
Interest in Using Tools and Machines	___	___	___	___	___	___	___	___	___	___
Interest in Rugged Activities	___	___	___	___	___	___	___	___	___	___
Interest in Protective Services	___	___	___	___	___	___	___	___	___	___
Interest in Technical Activities	___	___	___	___	___	___	___	___	___	___
Interest in Science	___	___	___	___	___	___	___	___	___	___
Interest in Leadership	___	___	___	___	___	___	___	___	___	___
Interest in Artistic Activities	___	___	___	___	___	___	___	___	___	___
Interest in Efficiency and Organization	___	___	___	___	___	___	___	___	___	___

INSTRUCTIONS FOR ATTRIBUTE RANKING: IMPORTANCE

In this task, you will be asked to rank order the thirty-one attributes according to their importance for *overall* performance in the MOS you supervise. In considering overall performance, think about all areas of job performance. You may want to review the performance definition and the attribute definitions before beginning.

Look at the list of thirty-one attributes. Think about how important each attribute is for overall performance in the MOS you supervise. Pick the attribute that you think is most important and place a "1" in the space beside the attribute. Next, pick the attribute that you think is least important and place a "31" beside the attribute. Then, pick the attribute that you think is second most important and place a "2" beside the attribute. Next, pick the attribute that you think is the second least important and place a "30" to the left of the attribute. Complete the rest of the rank orderings in this way, alternating back and forth from most to least important, until you have ranked all thirty-one attributes. When you are finished, review the rank orders to be certain that you have used each number (1 through 31) only once.

Please refer to the diagrams as you go through the following example.

A. Joe supervises 95B, Military Police. After reviewing the attribute definitions and the definition of job performance, Joe decides that Physical Strength is the most important attribute for overall performance as a 95B. So, Joe puts a "1" beside Physical Strength.

___ Verbal Ability
___ Reasoning
___ Number Ability
___ Spatial Ability
___ Closure
___ Mental Information Processing
___ Perceptual Speed and Accuracy
___ Memory
___ Mechanical Comprehension
___ Eye-Limb Coordination
___ Precision

___ Movement Judgment
___ Hand and Finger Dexterity
_1_ Physical Strength
___ Physical Endurance
___ Balance and Flexibility
___ Involvement in Athletics
___ Work Orientation
___ Sociability
___ Cooperation/Stability
___ Energy
___ Conscientiousness
___ Dominance/Confidence

___ Interest in Using Tools and Machines
___ Interest in Rugged Activities
___ Interest in Protective Services
___ Interest in Technical Activities
___ Interest in Science
___ Interest in Leadership
___ Interest in Artistic Activities
___ Interest in Efficiency and Organization

B. Next, Joe decides that Interest in Artistic Activities is least important for overall performance. So, he puts a "31" beside Interest in Artistic Activities.

___ Verbal Ability	___ Movement Judgment	___ Interest in Using Tools and Machines
___ Reasoning	___ Hand and Finger Dexterity	___ Interest in Rugged Activities
___ Number Ability	1 Physical Strength	___ Interest in Protective Services
___ Spatial Ability	___ Physical Endurance	___ Interest in Technical Activities
___ Closure	___ Balance and Flexibility	___ Interest in Science
___ Mental Information Processing	___ Involvement in Athletics	___ Interest in Leadership
___ Perceptual Speed and Accuracy	___ Work Orientation	31 Interest in Artistic Activities
___ Memory	___ Sociability	___ Interest in Efficiency and Organization
___ Mechanical Comprehension	___ Cooperation/Stability	
___ Eye-Limb Coordination	___ Energy	
___ Precision	___ Conscientiousness	
	___ Dominance/Confidence	

C. Joe then decides that Verbal Ability is the second most important attribute for overall performance. He puts a "2" beside Verbal Ability.

2 Verbal Ability	___ Movement Judgment	___ Interest in Using Tools and Machines
___ Reasoning	___ Hand and Finger Dexterity	___ Interest in Rugged Activities
___ Number Ability	1 Physical Strength	___ Interest in Protective Services
___ Spatial Ability	___ Physical Endurance	___ Interest in Technical Activities
___ Closure	___ Balance and Flexibility	___ Interest in Science
___ Mental Information Processing	___ Involvement in Athletics	___ Interest in Leadership
___ Perceptual Speed and Accuracy	___ Work Orientation	31 Interest in Artistic Activities
___ Memory	___ Sociability	___ Interest in Efficiency and Organization
___ Mechanical Comprehension	___ Cooperation/Stability	
___ Eye-Limb Coordination	___ Energy	
___ Precision	___ Conscientiousness	
	___ Dominance/Confidence	

D. Next, Joe decides that Closure is the <u>second least</u> important attribute, so he puts a "30" beside Closure.

2 Verbal Ability
__ Reasoning
__ Number Ability
__ Spatial Ability
30 Closure
__ Mental Information Processing
__ Perceptual Speed and Accuracy
__ Memory
__ Mechanical Comprehension
__ Eye-Limb Coordination
__ Precision

__ Movement Judgment
__ Hand and Finger Dexterity
1 Physical Strength
__ Physical Endurance
__ Balance and Flexibility
__ Involvement in Athletics
__ Work Orientation
__ Sociability
__ Cooperation/Stability
__ Energy
__ Conscientiousness
__ Dominance/Confidence

__ Interest in Using Tools and Machines
__ Interest in Rugged Activities
__ Interest in Protective Services
__ Interest in Technical Activities
__ Interest in Science
__ Interest in Leadership
31 Interest in Artistic Activities
__ Interest in Efficiency and Organization

Joe continues to rank order the attributes in this way, alternating back and forth between most and least important, until all thirty-one attributes have been rank-ordered. As a final check, Joe makes certain that he used each number from 1 through 31 only once.

ATTRIBUTE RANKING WORKSHEET

___ Verbal Ability

___ Reasoning

___ Number Ability

___ Spatial Ability

___ Closure

___ Mental Information Processing

___ Perceptual Speed and Accuracy

___ Memory

___ Mechanical Comprehension

___ Eye-Limb Coordination

___ Precision

___ Movement Judgment

___ Hand and Finger Dexterity

___ Physical Strength

___ Physical Endurance

___ Balance and Flexibility

___ Involvement in Athletics

___ Work Orientation

___ Sociability

___ Cooperation/Stability

___ Energy

___ Conscientiousness

___ Dominance/Confidence

___ Interest in Using Tools and Machines

___ Interest in Rugged Activities

___ Interest in Protective Services

___ Interest in Technical Activities

___ Interest in Science

___ Interest in Leadership

___ Interest in Artistic Activities

___ Interest in Efficiency and Organization

Please check to see that you used each number from 1 through 31 only once.

Rated MOS ___ Your Rank ___ Today's Date ______________

Method ______ Name ______________________________________

ATTRIBUTE MODEL EVALUATION SHEET

Performance Definition

Please recall (review, if necessary) the materials used to define performance in the MOS. Then, answer the following questions.

1. How would you describe the amount of detail and information provided in the performance definition?

1	2	3	4	5	6	7
Far Too Little			The Right Amount			Far Too Much

2. If you think the performance definition has too little detail and information, what would you add?

3. If you think the performance definition has too much detail and information, what would you delete?

Attribute Ranking

1. How clear were the instructions for the ranking?

1	2	3	4	5	6	7
Not At All Clear			Neither Clear Nor Unclear			Very Clear

2. Was it easy or difficult to rank order the attributes?

1	2	3	4	5	6	7
Very Difficult			Neither Easy Nor Difficult			Very Easy

3. How confident do you feel about the accuracy of your rankings?

1	2	3	4	5	6	7
Not At All Confident			Somewhat Confident			Very Confident

Attributes

1. Think about the list of thirty-one attributes (below). Place a checkmark in front of any attributes that were difficult to judge or rank. Please check at least two attributes.

___ Verbal Ability

___ Reasoning

___ Number Ability

___ Spatial Ability

___ Closure

___ Mental Information Processing

___ Perceptual Speed and Accuracy

___ Memory

___ Mechanical Comprehension

___ Eye-Limb Coordination

___ Precision

___ Movement Judgment

___ Hand and Finger Dexterity

___ Physical Strength

___ Physical Endurance

___ Balance and Flexibility

___ Involvement in Athletics

___ Work Orientation

___ Sociability

___ Cooperation/Stability

___ Energy

___ Conscientiousness

___ Dominance/Confidence

___ Interest in Using Tools and Machines

___ Interest in Rugged Activities

___ Interest in Protective Services

___ Interest in Technical Activities

___ Interest in Science

___ Interest in Leadership

___ Interest in Artistic Activities

___ Interest in Efficiency and Organization

2. This list of attributes was designed to cover all the attributes which are relevant to Army MOS.

   Are there any attributes you would add?

   Are there any attributes you would delete?

General Evaluation

Think about the two tasks you completed, attribute judgments and ranking. Then answer the following questions.

1. Which was easier?

   ___ Judgment

   ___ Ranking

2. Which provides more information about the MOS?

   ___ Judgment

   ___ Ranking

3. If the Army was going to use one of these methods to guide selection and placement of soldiers in MOS, which would you prefer?

   ___ Judgment

   ___ Ranking

APPENDIX E

DETAILED RESULTS OF TASK MATCHING EXERCISE

APPENDIX E: Mapping of Project A Tasks onto Important Task Categories

Categories with High Mean Importance (X > 3.00) 11B	63B	71L	Task Category	Tasks Matched (50% or Greater Agreement) 11B	63B	71L
*	*		1. Perform operator maintenance checks and service		10	
*	*		2. Perform operator checks and services on weapons	01,14	18,21	15,16
	*		3. Inspect mechanical systems			
			4. Repair weapons			
	*		5. Repair mechanical systems		01,04,05,05 06,07,14,15	
	*		6. Troubleshoot mechanical systems		11,13,14,27 28,29	
			7. Troubleshoot weapons			
*			8. Install electronic components			
	*		9. Inspect electrical systems			
			10. Inspect electronic systems			
	*		11. Repair electrical systems		03	
			12. Repair electronic components			
	*		13. Troubleshoot electrical systems		12	
			14. Troubleshoot electronic components			
			15. Pack and load materials			
			16. Prepare parachutes			
			17. Prepare equipment and supplies for air drop			
			18. Operate power excavating equipment			
*	*		19. Operate wheeled vehicles		09	
*			20. Operate track vehicles			
			21. Operate boats			
			22. Operate lifting, loading, and grading equipment			
			23. Paint			
			24. Install wire and cables			
			25. Repair plastic and fiberglass			

APPENDIX E: Mapping of Project A Tasks onto Important Task Categories

Categories with High Mean Importance (X > 3:00)				Tasks Matched (50% or Greater Agreement)		
11B	63B	71L	Task Category	11B	63B	71L
			26. Repair metal			
			27. Assemble steel structures			
			28. Install pipe assemblies			
			29. Construct wooden buildings and other structures			
			30. Construct masonry buildings and structures			
			31. Operate gas and electric powered equipment			
			32. Select, layout and clean medical equipment			
			33. Use audiovisual equipment			
		*	34. Reproduce printed material			
			35. Operate electronic equipment			
			36. Operate radar			
		*	37. Operate computer hardware			
			38. Cook			
			39. Perform medical laboratory procedures			
			40. Conduct land surveys			
			41. Provide medical or dental treatment			
*			42. Sketch maps, overlays, or range cards	16		
			43. Produce technical drawings			
			44. Draw maps and overlays			
			45. Draw illustrations			
		*	46. Type			01,02,03,05,06 10,11
		*	47. Prepare technical forms and documents			01,03,04
		*	48. Record, file, and dispatch information			07,08,09,12,13 14
		*	49. Receive, store, and issue supplies, equipment, and other materials			27,28
*			50. Use hand and arm signals			

APPENDIX E: Mapping of Project A Tasks onto Important Task Categories

Categories with High Mean Importance (X > 3.00)				Tasks Matched (50% or Greater Agreement)		
11B	63B	71L	Task Category	11B	63B	71L
*	*	*	51. Read technical manuals, field manuals, regulations, and other publications		01,03,04,05,06 07,08,10,11,12 13,14,15,18,27 28,29,30	08
*	*		52. Use maps	13,27	19	19
*			53. Send and receive radio messages	09		
*			54. Give oral reports	10		
			55. Receive clients, patients, guests			
*		*	56. Give directions and instructions			
		*	57. Write documents and correspondence			
			58. Write and deliver presentations			
			59. Interview			
			60. Provide counseling and other interventions			
*			61. Decode data			
			62. Analyze electronic signals			
			63. Analyze weather conditions			
			64. Order equipment and supplies			
			65. Estimate time and cost of maintenance operation			
			66. Plan placement or use of tactical equipment			
			67. Translate foreign languages			
			68. Analyze intelligence data			
			69. Control money			
			70. Determine firing data for indirect fire weapons	24		
			71. Compute statistics/mathematical calculations			
			72. Provide programming and data processing support support for compurter operations			
			73. Control air traffic			
*			74. Use hand grenades	18		

APPENDIX E: Mapping of Project A Tasks onto Important Task Categories

Categories with High Mean Importance (X > 3.00)				Tasks Matched (50% or Greater Agreement)		
11B	63B	71L	Task Category	11B	63B	71L
*	*	*	75. Protect against NBC hazards	23,30	23,24,16	22,23,24,26
*			76. Handle demolitions or mines	19		
*			77. Engage in hand-to-hand combat			
*	*	*	78. Fire individual weapons	07		
*			79. Control individuals and crowds			
*			80. Customs and laws of war			25
*			81. Navigate	27	20	19,20
*	*	*	82. Survive in the field			17,18
			83. Load and unload field artillery or tank guns			
			84. Fire heavy direct fire weapons (e.g., tank main guns, TOW, missile, IFV cannon)			
			85. Prepare heavy weapons for tactical use			
*			86. Place and camouflage tactical equipment and materials in the field	22	17	17
			87. Fire indirect fire weapons			
*	*	*	88. Give first aid	25,26	16,25	21
*			89. Detect and identify targets	11		
			90. Plan, organize, monitor			
			91. Clarify roles, provide feedback			
			92. Provide information			
			93. Recognize, reward			
			94. Support			
			95. Train, develop			
			96. Discipline, punish			
	*		97. Act as model			
			98. Conduct tactical operations			

APPENDIX E: Mapping of Project A Tasks onto Important Job Activities

Activities with High Mean Importance (X > 3.00) 11B	63B	71L	Job Activity	Tasks Matched (50% or Greater Agreement) 11B	63B	71L
*	*	*	1. Work in a team	02		
*			2. Lead a team			
			3. Support/counsel peers			
			4. Support/counsel subordinates			
*		*	5. Coach peers			
*			6. Coach subordinates			
*	*	*	7. Make oral reports (to individuals)			
			8. Make oral reports (to groups)			
*		*	9. Relay oral instructions			
			10. Ask questions			
		*	11. Record information		10	
			12. Write brief messages			
			13. Write longer reports			
			14. Monitor/interpret verbal messages			
*	*		15. Recall verbal information		22	
			16. Monitor/interpret numerical information			
			17. Recall numerical information			
	*		18. Monitor/interpret figural information			
			19. Recall figural information			
*	*	*	20. Follow oral directions			
*	*	*	21. Follow written directions		03,04,05,06,07 10,11,12,13,14 15,18,20,27,28 29,30,08	01,02,03,04,05 06,07,08,09,10 11,12,13,14,21 22,27
*			22. Judge size and distance			
*			23. Judge location			
			24. Judge paths of moving objects			

APPENDIX E: Mapping of Project A Tasks onto Important Job Activities

Activities with High Mean Importance (X > 3.00) 11B	63B	71L	Job Activity	Tasks Matched (50% or Greater Agreement) 11B	63B	71L
	*		25. Solve electrical system problems		03,12	
	*		26. Solve mechanical system problems		04,06,07,11,13 14,26,27,28,29 08	
			27. Solve logistical problems			
			28. Solve tactical maneuver problems			
		*	29. Solve administrative problems			
			30. Solve leadership problems			
			31. Solve medical problems		16	
			32. Solve communication problems			
	*		33. Operate precision hand-held equipment			
	*		34. Operate hand-held tools		04,05,06,07,08 15	
	*		35. Operate hand-held power equipment			
			36. Operate larger power equipment			
		*	37. Operate full keyboard			01,02,03,05,06
		*	38. Operate numeric keyboard			10,11,
	*		39. Adjust control device using one limb			
	*		40. Adjust control device multiple limbs			
*			41. Drive tracked vehicle			
			42. Drive heavy wheeled vehicle			
*	*		43. Drive light wheeled vehicle			
*	*		44. Aim: stationary target			
			45. Aim: moving target			
			46. Walk long distances			
			47. Run short distances			
	*		48. Push, pull, lift heavy weights			
			49. Throw objects	18		

APPENDIX E: Mapping of Project A Tasks onto Important Job Activities

Activities with High Mean Importance (X > 3.00)				Tasks Matched (50% or Greater Agreement)		
11B	63B	71L	Job Activity	11B	63B	71L
			50. Sort, fold, feed by hand			
	*		51. Make coordinated movements			
	*		52. Work long hours			
	*		53. Work under adverse conditions			

APPENDIX F

SUMMARY OF EVALUATION RATINGS BY METHOD, RATER TYPE, AND MOS

Summary of PERFDEF1 by Method, Rater Type, and MOS

Variable	Value Label	Mean	Std Dev	Cases
For Total Sample		4.55	.96	127
METHOD	Importance	4.58	.95	53
TYPE	Civilian	5.00	0.0	5
MOS	11B	5.00	0.0	1
MOS	63B	5.00	0.0	3
MOS	71L	5.00	0.0	1
TYPE	NCO	4.48	.95	35
MOS	11B	4.65	.93	20
MOS	63B	4.33	.58	3
MOS	71L	4.25	1.06	12
TYPE	Officer	4.69	1.11	13
MOS	11B	4.00	0.0	1
MOS	63B	5.00	1.41	5
MOS	71L	4.57	.98	7
METHOD	Validity	4.53	.98	74
TYPE	Civilian	4.50	.55	6
MOS	11B	4.50	.55	6
TYPE	NCO	4.60	1.22	40
MOS	11B	4.31	1.35	16
MOS	63B	4.65	1.11	17
MOS	71L	5.14	1.07	7
TYPE	Officer	4.43	.63	28
MOS	11B	4.57	.76	14
MOS	63B	4.25	.50	4
MOS	71L	4.30	.48	10

Summary of RANK1 by Method, Rater Type, and MOS

Variable	Value Label	Mean	Std Dev	Cases
For Total Sample		5.61	1.32	127
METHOD	Importance	6.17	.78	53
TYPE	Civilian	5.40	.55	5
MOS	11B	5.00	0.0	1
MOS	63B	5.67	.58	3
MOS	71L	5.00	0.0	1
TYPE	NCO	6.23	.81	35
MOS	11B	6.25	.79	20
MOS	63B	6.67	.58	3
MOS	71L	6.08	.90	12
TYPE	Officer	6.31	.63	13
MOS	11B	6.00	0.0	1
MOS	63B	6.80	.45	5
MOS	71L	6.00	.58	7
METHOD	Validity	5.22	1.47	74
TYPE	Civilian	4.33	1.86	6
MOS	11B	4.33	1.86	6
TYPE	NCO	4.92	1.58	40
MOS	11B	4.31	1.66	16
MOS	63B	5.41	1.37	17
MOS	71L	5.14	1.57	7
TYPE	Officer	5.82	.98	28
MOS	11B	5.57	1.16	14
MOS	63B	6.00	.82	4
MOS	71L	6.10	.74	10

Summary of RANK2 by Method, Rater Type, and MOS

Variable	Value Label	Mean	Std Dev	Cases
For Total Sample		4.27	1.59	127
METHOD	Importance	4.51	1.67	53
TYPE	Civilian	4.40	1.34	5
MOS	11B	6.00	0.0	1
MOS	63B	4.33	1.15	3
MOS	71L	3.00	0.0	1
TYPE	NCO	4.63	1.77	35
MOS	11B	4.75	1.74	20
MOS	63B	3.00	2.00	3
MOS	71L	4.83	1.70	12
TYPE	Officer	4.23	1.59	13
MOS	11B	5.00	0.0	1
MOS	63B	4.60	1.82	5
MOS	71L	3.86	1.57	7
METHOD	Validity	4.09	1.52	74
TYPE	Civilian	3.33	1.21	6
MOS	11B	3.33	1.21	6
TYPE	NCO	4.28	1.62	40
MOS	11B	4.25	1.53	16
MOS	63B	4.12	1.69	17
MOS	71L	4.71	1.80	7
TYPE	Officer	4.00	1.41	28
MOS	11B	4.00	1.47	14
MOS	63B	3.00	1.63	4
MOS	71L	4.40	1.17	10

Summary of RANK3 by Method, Rater Type, and MOS

Variable	Value Label	Mean	Std Dev	Cases
For Total Sample		4.92	1.31	127
METHOD	Importance	5.24	1.04	53
TYPE	Civilian	5.00	.71	5
MOS	11B	6.00	0.0	1
MOS	63B	4.67	.58	3
MOS	71L	5.00	0.0	1
TYPE	NCO	5.23	1.19	35
MOS	11B	5.15	1.04	20
MOS	63B	5.00	1.73	3
MOS	71L	5.42	1.38	12
TYPE	Officer	5.38	.65	13
MOS	11B	6.00	0.0	1
MOS	63B	5.60	.55	5
MOS	71L	5.14	.69	7
METHOD	Validity	4.69	1.44	74
TYPE	Civilian	4.83	1.17	6
MOS	11B	4.83	1.17	6
TYPE	NCO	4.60	1.55	40
MOS	11B	4.44	1.50	16
MOS	63B	4.53	1.50	17
MOS	71L	5.14	1.86	7
TYPE	Officer	4.78	1.37	28
MOS	11B	4.78	1.37	14
MOS	63B	4.00	2.16	4
MOS	71L	5.10	.99	10

Selection and Classification Technical Area Working Paper RS-WP-88-3

IMPROVING THE SELECTION OF MOS 11H TOW GUNNERS

Mark Y. Czarnolewski

February 1988

REVIEWED BY:

Clinton B. Walker
Acting Team Chief,
Predictor Development

APPROVED BY:

Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Cleared by:

Newell K.Eaton, Director

Manpower and Personnel Research Laboratory

U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Improving the Selection of MOS 11H TOW Gunners

Can we use a selection test to identify soldiers who will be good 11H TOW Gunners?

In Research Focus No. 1, we reported that computerized and paper-and-pencil tests developed by the Army's Project A predict M-1 gunnery in the Unit Conduct of Fire Trainer (UCOFT). In this issue, we present results of a project, conducted for the CG TRADOC, which shows that Project A tests predict measures of TOW gunnery skill at the end of AIT for MOS 11H trainees.

Background

Each MOS requires target tracking skills to varying degrees. This project looked at the continuous tracking behaviors -- i.e., the tracking of a target from the start of a simulated missile launch through the simulated hitting of the target -- required of 11H's by the M-70 simulator of the ground mount TOW2 launcher. The M-70 simulator score captures the likelihood that a soldier's continuous tracking behavior will result in his hitting a moving target. In this project, soldiers were tested on the Project A battery when they entered training at Fort Benning. Their M-70 scores were obtained from official records after they graduated from AIT ($n = 355$). Additional data, including GO/NO GO scores on 8 other 11H tasks, such as M-70 simulator assembly and range finding skills, were collected and analyzed for a subsample of 81 soldiers.

Results

The Project A psychomotor composite -- comprising 1-handed and 2-handed tracking tasks -- predicted M-70 scores for 345 soldiers passing AIT, $\underline{r} = .25$. The graph summarizes the extent that the Project A psychomotor composite improves prediction of TOW gunnery tracking performance over that

predicted by the AFQT--a standard measure of trainability. (The AFQT categories I and II represent higher scores and categories IIIB-IVA represent lower scores.) The AFQT does significantly predict TOW gunnery tracking, $\underline{r} = .12$; however, the power of the Project A psychomotor composite over the AFQT in predicting M-70 tracking performance is shown by:

- o The Project A Tracking Composite sorting out the better M-70 scorers within relatively homogeneous groupings of soldiers defined by the AFQT;
- o The Project A Tracking Composite identifying those soldiers in lower AFQT categories who would have higher M-70 scores than soldiers in the higher AFQT categories;
- o The top third of scorers on the Project A Tracking Composite have, on the average, M-70 scores that are 21 points above the average M-70 score; a similar cut based on the AFQT results in M-70 scores that are only 8 points above the average M-70 score.

Project A tests also predicted the detailed training data for the subsample of 11H's (n = 81). The psychomotor composite had significant correlations with M-70 scores, $\underline{r} = .43$, and with a composite of Number of Attempts to achieve a GO on eight AIT tasks $\underline{r} = .32$. (The tasks include important 11H tasks, such as assembling the TOW2, performing pre-operational checks on the TOW2, and using an anti-armor range card). Project A spatial tests -- Orientation and Maze -- were the most consistent predictors of M-70 performance for this subsample. Specifically, scores on Orientation significantly predict M-70 tracking, $\underline{r} = .33$, and the Number of Attempts

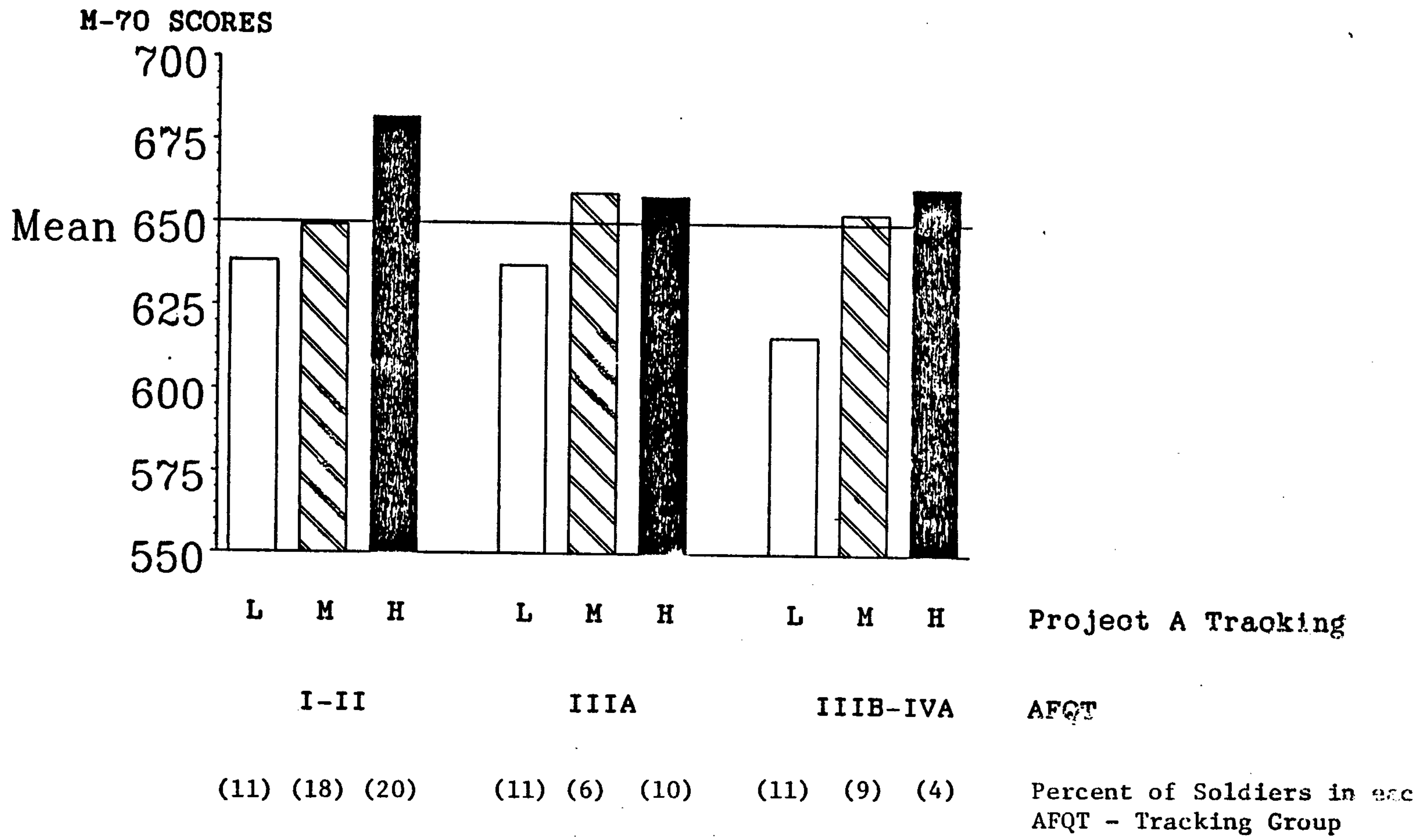
Composite, *r* = .34, and the Maze Test's significant correlations for these two criterion measures were, *r* = .32, and *r* = .38, respectively.

These results suggest that M-70 tracking performance and the other M-70 tasks require, at least, these skills:

- o Psychomotor -- 1-handed and 2-handed coordination for moving a marker to meet and remain on top of a target moving along a predefined path.
- o Spatial -- quickly scan for a target; determine its route; and, perhaps, maintain the target's referent point and maintain the reoriented position of the missile as it approaches the target.

The Project A Target Tracking, Orientation, and Maze Tests thus appear to reliably capture important components of 11H tasks and appear to provide additional power to the current selection process of choosing soldiers who would better perform those tasks.

MEAN M-70 SCORES FOR 11H's CLASSIFIED BY AFQT AND PROJECT A TRACKING MEASURES



NOTE - For Project A Tracking:
L = Lowest Third, M = Middle Third, H = Highest Third

Selection and Classification Technical Area Working Paper RS-WP-88-8

EXPLORATION OF VALIDITIES OF PROJECT A SPATIAL TESTS

Mark Y. Czarnolewski & Clessen J. Martin

September 1988

Approved for Distribution to Sponsor or Proponent.

REVIEWED BY:

Frances C. Grafton
Statistician

APPROVED BY:

Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Exploration of Validities of Project A Spatial Tests

The Joint-Service Selection and Classification Working Group (JSSCWG) provides technical and policy support to the Manpower Accession Policy (MAP) Steering Committee on matters related to ASVAB and other selection and classification issues. The Steering Committee is chaired by the Director for Accession Policy, Office of the Assistant Secretary of Defense (Force Management and Personnel) and is comprised of a senior military officer from the offices of the respective Service Deputy Chiefs of Staff for Personnel, and the Commander of the U.S. Military Entrance Processing Command (USMEPCOM). At the August, 1986 meeting of the Technical Task Group of the JSSCWG, each Service representative was requested to submit a candidate test of spatial ability at the November, 1986 JSSCWG meeting.

The technical representatives expressed special interest in the Army's recommendation because of the development of six new spatial tests for the Army's Project A: Improving the Selection, Classification and Utilization of Army Enlisted Personnel. Because considerable data had been collected on these tests during the concurrent validation phase of Project A, an informed decision concerning the selection of a spatial test could be made on the basis of the Project A data. The purpose of this report is to present the results of the validity analyses of the spatial tests in the nine MOS for which hands-on criterion measures had been developed.

METHOD

Subjects

A total of 5,196 first-term soldiers in nine military occupational specialties (MOS) provided the basis for the results of this report. Table 1 presents the nine MOS and the number of soldiers tested in each MOS.

Table 1

The Nine MOS and Sample Sizes Upon Which the Six Spatial Tests were Analyzed

MOS	DESCRIPTION	N
11B	INFANTRYMAN	694
13B	CANNON CREWMEMBER	653
19E	M48 - M60 ARMOR CREWMAN	499
31C	SINGLE CHANNEL RADIO OPERATOR	357
63B	LIGHT-WHEELED VEHICLE MECHANIC	630
64C	MOTOR TRANSPORT OPERATOR	678
71L	ADMINISTRATIVE SPECIALIST	504
91A	MEDICAL SPECIALIST	492
95B	MILITARY POLICE	689

Males comprised 88 percent of the sample. A total of 68 percent of the sample were white, 24 percent were black, three percent were Hispanic, four percent "other" and one percent indicated no racial origin.

Spatial Tests

Table 2 summarizes the psychometric characteristics of the six spatial tests. A brief verbal description of each test follows:

Assembling Objects. Each item requires the respondent to determine how an object will look when its parts are put together. Some items have the parts labelled with letters and require the respondent to match the letters, so

that the parts can be put together. Other items require the subject to mentally put shapes together to form one overall geometric shape.

Table 2

Summary Information[a] for the Six Spatial Tests Based in the Concurrent Validation Phase of Project A

Test Name	Construct	No. of Items	Testing Time	Split-Half Reliabilities[b]	Test-Retest Reliability[c]
Assembling Objects	Spatial Visualization	32	16 Min	.91	.70
Object Rotation	Spatial Visualization	90	7.5 Min	.99	.72
Maze Test	Spatial Visualization	24	5.5 Min	.90	.70
Orientation Test	Spatial Orientation	24	10 Min	.89	.70
Map Test	Spatial Orientation	20	12 Min	.90	.78
Reasoning Test	Induction	30	12 Min	.87	.65

[a]Information in this table is taken from Toquam, et al (1986)
[b]Odd-even procedure with Spearman-Brown correction based on Sample of 9332 to 9345.
[c]Based on samples sizes of 468 to 487.

Object Rotation. Each item consists of a test object and five response items. For each response item, the respondent must determine whether the response item matches the test item or not.

Maze Test. Each item is a rectangular maze with four labelled entrance points and four exit points. The task is to determine which of the four entrances leads through a pathway of the maze to one of the exit points.

Orientation Test. Each item contains a non-upright scene within a circular or rectangular frame. The frame has a circle with a dot inside of it. The

dot is placed inside the circle and at a point along the border of the circle. The task is to mentally rotate the frame to align the circle with the bottom of the scene. After doing so, the respondent must determine the new location of the dot inside the circle.

Map Test. Each item consists of a map that includes four landmarks spatially arranged to form a square. Within each item, the respondent is provided compass directions by indicating the direction of one landmark to another, for example, "the forest is North of the campsite." Respondents are also informed of their location relative to another landmark. Given this information, the respondent is to determine the direction of travel between two of the landmarks.

Spatial Reasoning Test. Each item consists of a series of four figures. The task is to identify the pattern or relationship among the four figures and then to identify among five possible answers the one figure that should appear next in the series.

Armed Forces Qualification Test (AFQT). The AFQT is a composite score of four ASVAB subtests - Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and Number Operations. The AFQT is used to classify soldiers and is included in the present report as a benchmark to provide an estimate of general ability level. The reliability of the AFQT is .92 or .93, depending on the ASVAB form that is used (Palmer, Hartke, Ree, Welsh, and Valentine, 1988).

Performance Criteria

Hands-on, Task Proficiency Tests. For each of the nine MOS, a set of 15 critical tasks of the task domain were selected for hands-on testing. Each task comprised several performance steps and each step was scored pass or

fail. A proportion-passed score was derived for each task and the proportions were averaged across tasks to yield an overall hands-on test score for each soldier.

School Knowledge Tests (K3). These written tests were developed to represent the instructional material developed within the Advanced Individual Training course for each of the nine MOS. Each test consists of 100 to 200 written items and covers 30 tasks.

Job Knowledge Tests (K5). These written tests consist of three to fifteen questions on each of a sample of 30 representative tasks and include the 15 tasks considered critical for hands-on testing. The K5 tests consist of 100 to 200 written job knowledge items.

In summary, the six Project A tests will be validated against job performance criteria for nine MOS. To the extent that the Project A spatial tests measure different abilities or skills, one may ask which skill best predicts performance and what are the unique characteristics of the test representing the skill. The focus of this paper is to concentrate on the unique characteristics of each spatial test and based on those characteristics decide on the "best" spatial ability test for the Army.

RESULTS

Table 3 presents the means and standard deviations for the Armed Forces Qualification Test (AFQT) and the Project A spatial tests for the nine MOS. Again, the AFQT is used to classify soldiers and is included as a benchmark to provide an estimate of general ability level. One would want to determine if this estimate of ability level corresponds with ability as defined by the other measures.

TABLE 3

Means and Standard Deviations for AFQT* and Mechanical Comprehension (MC) Scores from

ASVAB and Project A Spatial Measures for Nine MOS

MOS	n	AFQT[a]	Assembling Objects[b]	Map	Maze	Orientation	Object Rotation	Spatial Reasoning
11B	(439)	56.3 (22.0)	24.6 (6.1)	9.6 (5.5)	17.6 (4.3)	12.0 (6.2)	64.9 (18.9)	19.8 (5.7)
13B	(420)	47.0 (20.1)	22.7 (6.5)	7.0 (5.4)	16.2 (4.8)	10.1 (5.8)	60.5 (19.5)	18.4 (5.6)
19E	(357)	53.8 (21.6)	24.3 (5.9)	9.1 (5.7)	17.4 (4.5)	11.8 (6.3)	64.3 (18.7)	20.3 (5.5)
31C	(256)	59.1 (19.8)	24.2 (5.9)	8.5 (5.4)	17.3 (4.4)	11.9 (6.3)	64.3 (19.4)	20.5 (4.5)
63B	(429)	49.6 (19.3)	24.6 (6.0)	8.1 (5.5)	17.2 (4.4)	12.0 (6.3)	64.5 (19.0)	19.7 (5.5)
64C	(479)	42.9 (18.4)	23.0 (6.3)	7.0 (5.2)	16.3 (4.5)	10.3 (5.8)	62.5 (18.2)	18.4 (5.5)
71L	(389)	57.0 (19.3)	23.2 (6.6)	6.8 (5.1)	15.4 (4.6)	10.4 (5.8)	58.3 (17.8)	19.5 (5.2)
91A	(340)	61.0 (17.9)	25.8 (5.2)	8.9 (5.4)	17.0 (4.6)	12.8 (6.2)	64.5 (17.3)	21.2 (4.4)
95B	(499)	62.4 (17.0)	25.2 (5.3)	10.2 (5.4)	18.2 (3.9)	13.0 (6.0)	69.9 (15.6)	21.2 (4.3)

Note. ASVAB = Armed Services Vocational Aptitude Battery (See Text).
AFQT = Armed Forces Qualification Test Composite (See Text).
[a]The AFQT is a percentile score.
[b]Assembling Objects and the other spatial measures are number correct.

The 13B soldiers have the lowest average score on all the Project A spatial measures, except for the Maze Test, while the 95B soldiers are

consistently highest on all spatial measures. The 13B have the second lowest average AFQT and the 95B have the highest. These descriptive statistics indicate that the average spatial ability differences among the MOS coincide with the average AFQT differences among the MOS, and that in general, the Project A spatial measures are most likely affected by AFQT differences.

Table 4 presents the means and standard deviations for each MOS for the school knowledge (K3), job knowledge (K5), and hands-on measures. The K3, K5 and hands-on test scores are raw scores (i.e., percent correct). Table 4 also presents descriptive statistics for the Core Technical and General Soldiering tasks. The Core Technical tasks are composed of tasks unique to the MOS. The General Soldiering tasks are composed of tasks that are required of all soldiers, such as administering CPR, putting on clothing for Nuclear Biological Conflict (NBC), and preparing an M-16A rifle for firing. The Core Technical and General Soldiering scores are standardized scores that are a composite of K3, K5, and hands-on tests, as well as ratings of soldier performance.

Variables accounting for the average differences across the MOS hands-on and knowledge scores may include a number of characteristics related to soldier training. These characteristics may include (a) the Project A tasks are easier than other tasks in which soldiers are trained, (b) Project A tasks reflect the relatively lower sophistication required for some tasks in some MOS, or (c) Project A tasks represent tasks that are given much more practice, regardlessof their relative sophistication, with some MOS given more practice on tasks than are other MOS.

Table 4

<u>Means and Standard Deviations for School (K3), Job Knowledge (K5), Hands-on (HO), Core Technical and General Task Criterion Measures for Nine MOS</u>

	CRITERION MEASURES				
MOS:	K3	K5	HO	CORE TECH	GENERAL TASK
11B	62.1 (11.4)	61.9 (10.5)	70.4 (7.4)	102.9 (16.1)	102.9 (16.1)
13B	56.5 (10.3)	64.1 (10.2)	62.4 (11.3)	102.0 (16.9)	102.7 (15.4)
19E	68.7 (10.8)	63.3 (9.2)	77.8 (7.3)	102.7 (15.0)	102.5 (15.7)
31C	60.3 (11.4)	60.6 (9.9)	70.4 (7.8)	101.5 (17.0)	101.8 (16.6)
63B	61.0 (12.3)	66.2 (10.0)	84.7 (4.8)	102.6 (15.2)	102.3 (14.9)
64C	63.1 (10.6)	59.6 (9.4)	71.6 (7.6)	102.0 (14.5)	101.9 (16.3)
71L	59.2 (10.6)	58.2 (9.9)	59.3 (8.2)	101.2 (17.4)	101.3 (16.7)
91A	60.3 (10.3)	66.9 (9.2)	72.5 (7.0)	102.9 (15.9)	102.2 (16.4)
95B	59.4 (9.6)	62.9 (8.7)	70.4 (6.1)	100.9 (14.7)	101.1 (16.1)

*Core technical and general task are the same items for 11B.

Correlations of spatial ability measures with hands-on and other criterion measures may reflect the extent that these abilities predict job performance, despite the inherent variations across MOS. Where prediction of

a criterion measure (or measures) is consistent across MOS, one may ask, "What common characteristic is being predicted and how important is that characteristic for job competence?".

Tables 5 and 6 provide the correlations of the Project A spatial measures and the AFQT with the criterion measures. The correlations are the median correlations of each test with criterion measures across the nine MOS. The range of the correlations across the nine MOS are in parentheses. Table 5 contains correlations uncorrected for range restriction and Table 6 contains correlations corrected for range restriction. Table 5 answers the question, "Given the current system for selecting and classifying soldiers for the different MOS, which spatial test is the more consistent predictor of performance?". Table 6 answers the question, "Given the possibility of changing the selection and classification system and making the maximal use in using the spatial test for selecting and classifying soldiers, which spatial test could become the most consistent spatial predictor of performance?"

Table 5 shows that, even with a group of soldiers that are selected for the MOS and have stayed in the Army for 18-24 months, the Project A tests have moderate correlations with all the criterion measures, except for the hands-on measures. The AFQT is a similarly reasonable predictor. Table 6 shows that correcting for range restriction results in a change in the relative magnitude of the correlations of the spatial tests with the criterion measures. For example, the Map Test's median correlation with the K3 tests is 22 percent ($\underline{r} = .61$ vs. $\underline{r} = .50$) larger than the correlation of the Assembling Objects Test's median correlation when correcting for range restriction. When one does not correct for range restriction, the Map Test's

median correlation is only 14 percent (r = .40 vs. r = .35) larger than the Assembling Object's median correlation.

Table 5

Median and Range of Uncorrected Correlations Between Each Spatial Test and AFQT with School (K3), Job Knowledge (K5), Hands-On (HO), Core Technical and General Task Criterion Measures for the Nine MOS

	CRITERION MEASURES				
	K3	K5	HO	CORE TECH	GENERAL TASK
ASSEMBLING OBJECTS	35 (28-41)	43 (35-50)	20 (16-33)	34 (23-27)	37 (31-45)
MAP	40 (32-54)	44 (32-59)	24 (15-34)	31 (26-47)	38 (34-51)
MAZE	21 (15-26)	29 (20-33)	16 (15-27)	21 (10-25)	25 (15-30)
OBJECT ROTATION	17 (04-25)	22 (13-31)	19 (10-28)	19 (07-30)	23 (19-29)
ORIENTATION	31 (18-37)	35 (27-43)	22 (15-33)	28 (19-38)	33 (28-39)
SPATIAL REASONING	34 (27-43)	40 (36-51)	27 (13-38)	33 (23-44)	38 (30-46)
AFQT	40 (31-52)	40 (33-52)	21 (08-33)	35 (23-44)	37 (30-45)

This differential effect on the correlations when correcting for range restriction also results in a change in the rank of the spatial tests , when ranking them from best to worst predictors. For example, when correcting for

range restriction, the Spatial Reasoning Test has a higher correlation with K3 tests than does the Assembling Objects Test; without correcting for range

TABLE 6

Median and Range of Corrected Correlations Between Each Spatial Test and AFQT with School (K3), Job Knowledge (K5), Hands-On (HO), Core Technical and General Task Criterion Measures for the Nine MOS

	CRITERION MEASURES				
	K3	K5	HO	CORE TECH	GENERAL TASK
ASSEMBLING OBJECTS	50 (43-56)	58 (48-61)	34 (20-44)	46 (35-51)	52 (44-55)
MAP	61 (56-68)	64 (53-73)	41 (20-53)	53 (38-65)	58 (51-68)
MAZE	37 (32-49)	46 (39-51)	32 (17-43)	38 (29-48)	42 (36-51)
OBJECT ROTATION	35 (27-46)	37 (31-46)	27 (15-36)	35 (22-42)	38 (35-42)
ORIENTATION	48 (42-56)	53 (43-58)	36 (16-46)	45 (32-53)	49 (40-57)
SPATIAL REASONING	57 (42-60)	60 (49-65)	44 (18-52)	52 (31-60)	56 (43-64)
AFQT	67 (51-75)	66 (54-71)	42 (10-62)	61 (33-71)	63 (48-66)

restriction, the Assembling Objects Test has the higher correlation. A general statement would be that the tests with the largest validities, in order from highest to lowest, appear to be the Map, Spatial Reasoning,

Assembling Objects, and Orientation tests. Again, the AFQT appears to be a very good predictor of performance.

Two other criteria for determining the "best" spatial test may include (a) the extent of convergence of the Project A spatial tests with a previously used ASVAB spatial test and (b) the extent of divergence of the Project A spatial tests from the AFQT.

The ASVAB 5/6/7 spatial test was called the Space Perception Test. The test required that a respondent identify a three-dimensional figure obtained from folding a flat figure (Maier and Grafton, 1981). It had 20 items and required 12 minutes for administration. Test - retest reliabilities were .71 (Bayroff and Fuchs, 1970). Table 7 shows that the Space Perception Test correlates highest with the Assembling Objects Test, $\underline{r}$ = .41, and lowest with the Map and Spatial Reasoning Tests $\underline{r}$ = .28 ($\underline{N}$ = 431). To the extent that the Space Perception Test represents an important ability predictive of military performance, one may state that the Assembling Objects Test has more in common with that ability than the other Project A measures.

Alternatively, Table 7 shows that the Map Test has the highest correlations with the current AFQT, with the correlations between Assembling Objects and the AFQT being much lower. To the extent that the current AFQT is used as a standard for placing a soldier on a scale of trainability, one sees that Assembling Objects could potentially provide more information toward making that scale a better, more comprehensive measure of trainability than would the Map or the Spatial Reasoning Test.

Table 7

Correlations of ASVAB 5/6/7 Space Perception Test and the AFQT with Project A Spatial Tests

PROJECT A TESTS	ASVAB 5/6/7 SPACE PERCEPTION	AFQT MEDIAN	AFQT RANGE
ASSEMBLING OBJECTS	.41	.34	.30-.39
MAP	.28	.50	.46-.62
MAZE	.37	.30	.23-.34
OBJECT ROTATION	.26	.23	.18-.30
ORIENTATION	.31	.38	.33-.44
SPATIAL REASONING	.28	.48	.40-.52

Note. N = 431

[a]These correlations represent the median correlation and range of correlations of the AFQT with the Project A tests across the nine MOS. Sample sizes were too small for some MOS for computing the median correlation with the ASVAB 5/6/7 Space Perception Test.

CONCLUSION

One may conclude that the Project A spatial tests correlate with job performance reasonably well, although the spatial tests' correlations with hands-on performance appear to be relatively lower than with the other criterion measures. Four spatial tests appear to have good criterion validities, with the Map Test showing the strongest validities across the criterion measures. However, the Assembling Objects Test, which shows reasonable, although not as strong validities as the Map Test, does show a much higher convergence with a previous ASVAB spatial test. The Assembling

Objects Test also shows much more divergence from the AFQT, a composite of four ASVAB tests that is currently used to represent trainability potential. Given these combination of factors, it is concluded that the Assembling Objects Test shows more promise as a test that could add most toward constructing a measure of trainability (i.e., AFQT) that includes a spatial test.

REFERENCES

Bayroff, A.G., & Fuchs, E.F. (1970). The Armed Services Vocational Aptitude Battery. (Technical Research Report 1161). Alexandria, VA: Army Research Institute of the Behavioral and Social Sciences.

Maier, M.H., & Grafton, F.C. (1981). Aptitude Composites for ASVAB 8, 9, 10. (Research Report 1308). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Palmer, P., Hartke, D.D., Ree, M.J., Welsh, J.R., Valentine, L.D. (1988). Armed Services Vocational Aptitude Battery (ASVAB): Alternate Forms Reliability (Forms 8, 9, 10, and 11). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Toquam, J., Peterson, N. Rosse, R., Ashworth, S. Hanson, M.A., Hallam, G. (1986). Concurrent Validity Data Analyses: Cognitive Paper - and - Pencil and Computer - Administered Predictors (Trial Battery). Project A IPR & SAG Meeting. Minneapolis, Min. Army Research Institute of the Behavioral and Social Sciences.

Acknowledgment

The authors would like to acknowledge the technical support of Winnie Young and Lauress Wise for computing correlations most notably for the correlations corrected for range restriction, and to Frances Grafton for other data analysis and for substantial constructive feedback.

Working Paper

RS-WP-84-08

ALTERNATE METHODS OF ESTIMATING THE DOLLAR VALUE OF PERFORMANCE

Newell K. Eaton, Hilda Wing & Karen Mitchell
SELECTION AND CLASSIFICATION TECHNICAL AREA

February 1984

Reviewed by
Lawrence M. Hanser
Chief
Selection and Classification
Technology Transfer

Approved by
Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

PB 67335

Alternate Methods of Estimating the Dollar Value of Performance

Newell K. Eaton, Hilda Wing, and Karen J. Mitchell

U. S. Army Research Institute for the Behavioral and Social Sciences

Alexandria, Virginia

Running head: Dollar Value of Performance

Abstract

Techniques for estimating the dollar benefit of increased levels of productivity or improved performance call for estimation of the dollar value of performance at given levels and of SD$, the standard deviation of performance quality measured in dollars. Such estimates have been obtained via cost accounting or by supervisor estimates. This research attempts to estimate utility for various performance levels in the context of expensive, complex weapons systems. Changes in system performance are discussed, here, in terms of changes in the number of system units required for a fixed level of accomplishment. Formulae are presented for linking these changes to changes in the utility of the system in dollars. Information about system performance changes can also be entered into equations to compute SD$. These techniques for estimating the value of performance levels and determing SD$ may be useful additions to traditional cost accounting and newer supervisory estimation techniques.

Alternate Methods of Estimating the Dollar Value of Performance

Testing to improve selection and classification decisions is a normal part of entering into employment in most large organizations and in many small ones (Friedman & Williams, 1982). The benefits of such testing have been questioned on an increasingly frequent basis, both from the perspective of fairness and in terms of the benefits accrued from sometimes costly selection programs. Cost benefit analyses of selection and classification procedures frequently stress the price of recruiting and testing, while neglecting the dollar benefits of resulting increases in productivity. This oversight leaves managers unable to estimate the dollar value to the organization of selection and classification procedures.

The utility of implementing various personnel policies is important data for management decision-making. The costs of implementing personnel policies must be weighed against anticipated benefits. Implementation costs are usually couched in real dollar terms. Easily estimated costs are associated with salaries, space, overhead for test administration, fees, per diem paid to applicants, computer time and personnel for scoring, etc. Benefits accrued from implementation of personnel policies, however, are not as clearly identifiable in dollar terms. Judgments of the net positive impact of implementing given personnel policies are, therefore, difficult to make.

Early approaches to estimating the benefits associated with testing programs focused on the correlation between predictor measures and criterion performance. Kelly (1923) defined the "Index of Forecasting Efficiency" as $E = 1 - \sqrt{1 - r^2}$. A second index, the "Coefficient of Determination" was proposed by Hull (1928). This index is simply r^2, the proportion of criterion variance accounted for by the predictor measures. Both approaches suggest that only testing programs with high validity coefficients are beneficial.

However, important management information is neglected by mathematical procedures involving only validity or relative variance. Taylor and Russell (1939) incorporated information about selection ratios, the proportion of applicants selected to those who apply, and information about the base rate of successful criterion performance. Their computations dichotomize criterion performance into satisfactory and unsatisfactory, and clearly show that with sufficiently small selection ratios, even tests with low validities can lead to important increases in the numbers of successful performers. Increases in the proportion of successful performers predicted by the selection instrument can be linked to the increased usefulness or utility of an instrument.

Brogden (1949) and Cronbach and Gleser (1965) also addressed utility estimation. Their formulations dealt with continuous levels of criterion performance rather than the dichotomy used by Taylor and Russell (1939). They linked normally distributed performance levels to the values estimated for those performance levels, in dollars. A useful

formula for the gain in productivity, or utility (U$), obtained by using valid selection procedures includes (a) Ns, the number of individuals selected (b) SD$, the standard deviation of performance, scaled in a utility metric such as dollars; and (c) the average performance expected on the criterion by the selected group as estimated from a valid predictor, given by Rxy Zx:

$$U\$ = Ns\ SD\$\ Rxy\ Zx \quad (1)$$

To account for the cost (Ct) of testing each of Na applicants the formula was adjusted to

$$U\$ = Ns\ SD\$\ Rxy\ \ Zx - Na\ Ct$$

A more complete description of such formulations can be found in Cronbach and Gleser (1965), Hunter and Schmidt (1982), and Cascio (1982).

While the values of most of the variables on the right hand side of the Brogden-Cronbach-Gleser formulas are known, the estimation of SD$, the standard deviation of performance scaled in dollars, is problematic. A recent review by Hunter and Schmidt (1982) reports that only two published efforts have attempted the computation of SD$ using cost accounting methods.

An alternative to cost accounting estimates is to estimate the dollar values to the organization of performance at the 50th percentile level, the 85th percentile level (one standard deviation above the mean), and sometimes, the 15th percentile level (one standard deviation below the mean). The dollar difference between the 15% and 50%, and 50%

and 85% points provides an estimate of SD$. This "SD$ Estimation Technique" was used by Cascio and Silbey (1979) with second level managers in food and beverage sales (Mean = $30,000, SD$ = $9,500); by Schmidt, Hunter, McKenzie, and Muldrow (1979) with computer programmers (SD$ = $10,413); by Hunter and Schmidt (1982) with budget analysts (SD$ = $11,327); and by Bobko, Karren, and Parkington (1983) with insurance counselors (Median = $96,000, SD$ = $56,950). In the last study, actual sales data were also available and yielded sales-based statistics which were close to those obtained by the rating method (Median = $117,300, SD$ = $52,308).

The SD$ estimations reported above were derived in contexts where performance was easily measurable in dollars. The SD$ estimation questions developed by Hunter and Schmidt (1982) asked for estimates of the "value to the agency" of various performance levels. The questions were preceeded by instructions to "consider the cost of having an outside firm provide these products and services" (Schmidt et al., 1979, for computer programmers), or to "consider what the cost would be of having an outside consulting firm produce these products and services" (Hunter & Schmidt, 1982, for budget analysts). Both Cascio and Silbey (1979) and Bobko et al (1983) framed questions in terms of performance value and estimates of total yearly dollar sales.

It occurred to us that there may be situations where such estimates would be impractical. These could occur where the nature of the work was such that managers were more accustomed to considering the relative productivity of employees or crews than the costs of hiring

replacements. These could also occur where employees operate very complex, expensive equipment and/or are focal to the productivity of a costly system.

Two methods came to mind which might better serve in such circumstances. The first was somewhat like the SD$ Estimation Technique. Instead of using estimates of the dollar value of 85th percentile performance, however, the technique uses estimates of how many superior (85th percentile) performers would be needed to produce the output of a fixed number of average (50th percentile) performers. This estimate, combined with an estimate of the dollar value of average performance, provides an estimate of SD$, and is the basis for the name "Superior Equivalents Technique".

The second technique is a slight extension of the Brogden-Cronbach-Gleser formula. The derivation is based on a system of productive units, each unit with two possible output levels, average and superior. Overall system performance can be increased by increasing the performance of the units from average to superior, or by increasing the number of units with average performance. SD$ and the utility of the selection system can be indexed by the cost of the increased number of average performing units required to achieve the overall increase in system performance. Cost accounting procedures can be used to estimate the cost of each unit. We called this the "System Effectiveness Technique".

The purpose of this research was to develop the "Superior Equivalents Technique" and the "System Effectiveness Technique", and to apply both and the SD$ Estimation Technique to an existing system. We

chose a system where the equipment is complex and expensive and where operators and supervisors are far more accustomed to thinking about the value of performance levels in terms of operational output rather than dollar value. This system is comprised of the tanks and their crews in the U.S. Army.

Method

First, the Superior Equivalents and Systems Effectiveness Techniques were developed to the extent that they could be applied to tank units. Second, estimates required by the Superior Equivalents Technique and the more conventional SD$ estimation technique were combined into one questionnaire and administered to a group of tank commanders. Data required for the Systems Effectiveness Technique were obtained from the literature.

Superior Equivalents Technique Development

The Superior Equivalents Technique is conceptually similar to the conventional SD$ Estimation Technique which estimates SD$ from the difference in dollar estimates of the value of 85th and 50th percentile performance. The basic concept of the Superior Equivalents Technique is that of estimating the standard deviation of performance in performance units, and then converting to dollar units. We assumed that supervisors are accustomed to evaluating the relative performance of their employees, and can make accurate judgments in these terms. Accordingly, they should have little difficulty in estimating the number (N85) of 85th percentile employees required to equal the performance of some fixed number (N50) of average performers. Where the value of

average performance (V50) is known, or can be estimated, SD$ may be estimated by using the ratio of N50/N85 times V50 to obtain V85, and then subtracting V50. This reduces to:

$$SD\$ = V85 - V50 \qquad (1)$$

$$\text{but} \quad V85 = \frac{(V50)(N50)}{N85}$$

Hence,

$$SD\$ = \frac{(V50)(N50)}{N85} - 1 \qquad (2)$$

In our case we set N50 = 17 as a fixed number of tanks with average commanders, because there are 17 tanks in a tank company. We used two methods for estimating V50. First, we used an estimate based on the salary and benefits paid the average tank commander. In general, one might assume organizations pay average employees about what they are worth. Second, we used an item on our questionnaire asking the dollar value of average performance.

System Effectiveness Technique Development

The Systems Effectiveness Technique does not rely on estimates of SD$ based on differences in the estimated value of various performance levels. Instead, it is based on the concept that improved performance with an existing number of units (employee/machines) in a system would yield overall system output equal to that given existing performance per unit but with an increased number of units. The value of the improved performance would equal the cost of the increased number of units. When

the Brogden-Cronbach-Glaser formula is recast in these terms, the resulting extension provides a utility formula in which SD$ is replaced by more readily obtained cost and performance terms. Our derivations follow:

Let the cost of a single unit in a system be Cu. Let the overall systems effectiveness be Y. This may be achieved with varying numbers of units depending on the performance of the units. Or

$$Y = n1Y1 = n2Y2 \quad . \; . \; . \; niYi \tag{3}$$

where ni = number of units at performance level i, on a ratio scale, and Yi = mean performance of units at level i, on a ratio scale.

Examples of performance scales useable in this formula are probability of hits per firing (Army tank commander), number of convictions per year (detective), number of pupils achieving given a standard (teacher), or other frequency-type variables. In a system where improved performance Y2 is obtained from each of the initial n1 units, the overall improvement in system performance is

$$n1Y2 - n1Y1 = n1\ (Y2 - Y1) \tag{4}$$

The number of extra units operating at the initial performance level which are needed to achieve this improved performance is

$$\text{DEL } n = \frac{n1\ (Y2 - Y1)}{Y1} \tag{5}$$

The dollar value of improved performance is equivalent to the extra number of lower performing units needed times the cost per unit

$$U\$ = \underline{Cu\ DEL\ n}$$

or

$$U\$ = \frac{Cu\ n1\ (Y2 - Y1)}{Y1} \qquad (6)$$

Simply stated, the value in dollars of achieving an overall improvement in performance in a system equals the cost of adding the number of units required to effect the improvement where those added units operate at the initial performance level.

Estimating U$ using SD in performance units. The basic Brogden-Cronbach-Glaser formula works in any metric (m); U and SD need not be expressed in dollars. The overall improvement in system performance is

$$\underline{Um = Ns\ SDm\ Rxy\ Zx}$$

In output units of performance, Y, this is equal to (4)

$$\underline{n1\ (Y2 - Y1) = Ns\ SDy\ Rxy\ Zx}$$

Substituting into (6)

$$U\$ = \frac{Cu\ Ns\ SDy\ Rxy\ Zx}{Y1} \qquad (7)$$

Formula (7) more conveniently describes the utility in dollars of selection. This formula uses SDy, the standard deviation in output units of performance, rather than SD$, the standard deviation of performance in dollars.

Estimating SD$ using cost and performance data. Setting (1) and (7) equal

$$U\$ = Ns\ SD\$\ Rxy\ Zx = \frac{Cu\ Ns\ SDy\ Rxy\ Zx}{Y1}$$

and solving for SD$

$$SD\$ = \frac{Cu\ SDy}{Yl} \tag{8}$$

Or, SD$ equals the cost per unit times the ratio of the SD of performance to the initial mean level of performance. This proportion may be easily estimated in many situations. It is interesting to note that this parallels the Hunter & Schmidt (1982) notion that SD$ may be linked to some percentage of salary. Here, Cu is the cost of the unit in the system - equipment, support, and personnel - rather than salary. However, estimates from both (7) and (8) are appropriate only when the performance of the units in the system is to a major extent a function solely of the performance of the individual in the job under investigation. To the extent that it is not, corrections to these formulae would be required.

Instrument

A questionnaire was developed to obtain estimates of the dollar value of average (50th percentile) and superior (85th percentile) tank commander performance, and the estimated number of tanks with superior tank commanders needed to equal the performance of a standard company of 17 tanks with average tank commanders. Additional items were used to determine the confidence of respondents in their answers. Dollar value and number-of-tank items were fill-in-the-blank while confidence estimates were multiple choice (within $1,000, within $10,000, within $100,000, and other; or exact, within 1 tank, within 2-3 tanks, within 4-5 tanks, or more than 5 tanks).

Respondents

The questionnaire was administered to 48 M1 tank commanders enrolled in advanced training at a Continental U.S. Army post and to five other enrollees (one M48 Armor crewman and four U.S. Marine tank commanders). All were male: their racial/ethnic identification of Ss was not obtained. Their median number of years experience as a tank crew member was nine. It can be assumed that the sample was representative of the corps of tank commanders in the continental U.S.

Other data

Both the Superior Equivalents Technique and the System Effectiveness Technique required information from sources other than the questionnaire. To obtain another value of average performance for the Superior Equivalents Technique, we used published pay and allowance tables. In 1983 the base pay for Army enlisted personnel with ten years of service at the ranks expected for tank commanders ranged from $15,500 to $17,000. Non-taxable allowances for such items as housing could amount to more than $8,000 for the typical married tank commander with dependents. An estimate of an equivalent civilian salary would be about $30,000 per year.

Data for the System Effectiveness Technique were obtained from technical reports of previous research. The cost of operating a tank per year, Cu, was estimated by Wallace (Note 2) as $3,000,000. This amount appears not to include the 20+ year life span expected for an Army tank; we preferred the more conservative figure of $300,000 per

year. Criterion related validity research on tank crew performance (e.g., Eaton, Bessemer & Christiansen, 1981) suggested that meaningful values for the ratio SDy/Yl range from .2 to .5. We selected the more conservative value of .2.

To permit computation of U$ for purposes of example, it was necessary to identify a selection ratio and a selection procedure validity. Inspection of tank doctrine indicated tank commanders could be chosen from tank drivers, gunners, or loaders. However, in practice only more senior crew members are considered. We chose .5 as most likely to reflect actual selection ratios. Although higher validities are often observed (Eaton, Note 1 and 1980) a conservative value of RXY = .3 was chosen.

Data from the questionnaire and the 'other data' were assembled to provide the basic input to each of the three techniques. Then, for each technique, SD$ was computed, and U$ was determined on an individual tank/tank commander basis, as well as for a system having 2500 tanks and tank commanders.

Results

SD$ Estimation Technique

The estimates of value for both average and superior tank commanders (TCs) were skewed and very broad. For the average TC, the median estimate was $30,000, with the first quartile at $17,000 and the third quartile at $100,000. The distribution appeared somewhat bimodal, with nineteen judges giving estimates between $10,001 and $30,000, and

sixteen judges giving estimates over $75,000. The mean of the distribution was $340,198 with a standard deviation of $952,798. The distribution is shown in Figure 1.

Insert Figure 1 about here

For the superior TC, the median estimate was $50,000 with the first quartile at $30,000 and the third quarter at $300,000. This distribution was also highly skewed and almost trimodal. There were nine estimates of $1,000,000 or more, including one at positive infinity which we set equal to $100,000,000. The distribution is shown in Figure 2. The mean is $2,899,806 and standard deviation is $14,071,535. Most (70%) of the subjects indicated their estimates for both average and superior TCs, were accurate within $10,000, with 11% not responding, and 19% indicating confidence limits greater than $10,000.

Insert Figure 2 about here

The difference between the median estimates of superior and average performance value provided an estimate of <u>SD$</u> = $20,000. The difference between average and superior performance value estimates given by each respondent was also computed. The median of the differences was 15,000, with the first quartile being $7,000 and the third, $85,000.

For a selection ratio of .5, <u>Zx</u> = .8 (Hunter and Schmidt, 1982).

Incorporating $\underline{r} = .3$, $\underline{Zx} = .8$, a <u>SD$</u> of $20,000 into (1) yielded <u>U$</u> = $4800 if one tank commander were selected, and <u>U$</u> = $12,000,000 if 2500 tank commanders were selected. With the <u>SD$</u> estimate of $15,000, the <u>U$</u> values are $3600 and $9,000,000 respectively. Values for <u>SD$</u> and <u>U$</u> are shown in Table 1.

Insert Table 1 about here

<u>Superior Equivalents Technique</u>

The modal response given for the number of superior TCs judged equivalent to 17 average TCs was 10. The median estimate was 9 and the median confidence was that the estimate was in error by no more than 1. The range of 4 to 12 included about 95% of the responses. About 85% said their estimate was in error by no more than 3. The response "10" was judged to be a representative value of central tendency. The distribution is shown in Figure 3.

Insert Figure 3 about here

The estimated value of an average tank commander was $30,000 from the <u>SD$</u> Estimation Technique, and also from evaluation of tank commander's pay and allowances. Given 10 superior TCs judged equivalent to 17 average TCs, and an average TC "worth" $30,000 per year, then a superior TC would be "worth" 17/10 times $30,000, or $51,000. This figure is almost the same as the median estimate of the value of a

superior tank commander obtained with the SD$ Estimation Technique. Values for U$ were obtained for 1 and 2500 tank commanders selected, yielding values of $5040 and $12,600,000 respectively. These values are also shown in Table 1.

System Effectiveness Technique

The value of SD$ was computed directly from formula (8) using the values Cu = $300,000 and SDy/Y1= .2, yielding SD$ = $60,000. Values of U$ for 1 and 2,500 tank commander selections were computed as $14,400 and $36,000,000 respectiely. These are shown in Table 1.

Discussion

The various methods of estimating SD$ and U$ do not all lead to the same results. Although raters providing estimates for the SD$ estimation technique indicated confidence in their judgments, the distribution of responses is far from comforting. The extreme skewing of judgments and their bi/tri- modal nature suggested that all responses were not based on the same considerations and values. This supports our concern about the difficulty of making such judgments when the cost of contracting out the work is unknown and/or other financially intangible factors are involved. Such is frequently the case for public employees such as the military when private industry counterparts are nonexistent.

Some confirmatory evidence for the SD$ estimation results comes from the similarity of results obtained using the Superior Equivalents Technique. The value of average performance, $30,000, was the same regardless of whether obtained from the instrument (used in the SD$ Estimation Technique) or from the 'independent' pay and allowances

information. And the ratio of 10 tanks with superior commanders equaling 17 with average commanders produced a SD$ of $21,000, practically equal to the $20,000 figure from the SD$ Estimation Technique.

Both techniques yielded similar results, and both are within the range of 40-70% of annual salary as noted by Hunter and Schmidt (1982) and (1983) to be typical. We believe, however, that the Superior Equivalents Technique is the method of choice in situations where supervisors are more accustomed to dealing with performance in output terms, or in relative output of individuals, than in dollar terms. In such situations, supervisor's judgments of relative performance seem more credible than their estimates of the dollar value of average and superior performance. The relative variances of the distributions of these values in this research would seem to support this point.

Despite their consistency, we wonder whether both SD$ estimation and superior equivalents methods produce underestimates. Perhaps respondents estimates are pay/allowance-based, and these elements are not indexed to the tank commander job in terms of 'real worth.' Respondent's estimates of the value of average performance were obtained separately from our evaluation of pay and allowances, but, of course, tank commanders know well their remuneration. Many respondents may have judged the performance of an average tank commander to be worth what he is paid.

Higher values for SD$ and U$ result from the Systems Technique. These are probably as accurate as any estimates. They were based on

SDy/Yl ratios from actual field data and cost estimates which, while crude, may more accurately reflect reality. It is also plausible that these Systems Cost Estimates are below the "true" values. A tank probably costs more than $300,000 per year to operate, if not the $3,000,000 proposed by Wallace (Note 2); and .2 is the lower bound for the ratio of SDy to Yl.

One could argue that improved performance of tank commanders, the basis for the SD$ Estimation and Superior Equivalents Techniques, has only a partial impact on improving tank performance, the basis of the system effectiveness procedures. An empirical question is the size of the contribution of the tank commander to the crew and tank; we do not yet know the answer. However, both analytical and rational judgments, as well as the lore within the armor community, suggest that the performance of the tank is largely a function of the performance of the tank commander. It was the assumption of the overwhelming impact of the tank commander on tank performance that led to our initial thoughts on the Systems Effectiveness formulae derivations.

The two additional limitations we see of the Systems Cost Technique are, first, the problem of estimating the cost of a unit in the system and, second, its restriction to those situations in which the quality of performance can be indexed by quantity. Wallace's (Note 2) estimate of unit cost per tank (Cu) may seem excessive. Nevertheless, such estimates can be, and frequently are, made in accounting departments. It also can be adjusted if it appears out of line. Second, quality as

indexed by quantity may not be meaningful in some situations. Prospective users should question, however, whether qualitative indices may not be made into quantitative ones. For example, a police department may decide that conviction of one murderer is equivalent to the conviction of several burglars. It is likely that such equations are being used to compare performance of different individuals, albeit informally and, possibly, inconsistently.

Reference Notes

1. Eaton, N. K. Predicting Tank Gunnery Performance (Research Memorandum 78-6, NTIC-ADA 077955). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, February, 1978.

2. Wallace, J. R. The Gideon Criterion: The effects of selection criteria on soldier capabilities and battle results. (Research Memorandum 82-1). Fort Sheridan, IL: U.S. Army Recruiting Command, January 1982.

References

Arnold, J. D., Rauschenberger, J. M., Soubel, W. G. & Guion, R. M. Validation and utility of a strength test for selecting steelworkers. Journal of Applied Psychology, 1982, 67, 5, 588-604.

Bobko, P., Karren, R., & Parkington, J. J. The estimation of standard deviations in utility analyses: An empirical test. Journal of Applied Psychology, 1983, 68, 170-176.

Brogden, H. E. When testing pays off. Personnel Psychology, 1949, 2, 171-183.

Cascio, Wayne F. Costing Human Resources: The Financial Impact of Behavior in Organizations. Boston: Kent Publishing Co., 1982.

Cascio, W. F. & Silbey, V. Utility of the assessment center as selection device. Journal of Applied Psychology, 1979, 64, 2, 107-118.

Cronbach, L. J. & Gleser, G. C. Psychological Tests and Personnel Decisions. (Second edition) Urbana: University of Illinois Press, 1965.

Eaton, N. K. Performance motivation in armor training. JSAS Catalogue of Selected Documents in Psychology, 1980, 10, 28.

Eaton, N. K., Bessemer, D. W., & Kristiansen, D. M. Tank crew position assignment. JSAS Catalogue of Selected Documents in Psychology, 1981, 11, 62-63.

Friedman, T. & Williams, E. B. Current use of tests for employment. Ability testing: Uses, consequences and controversies. Part II: Documentation section. Washington, DC: National Academy Press, 1982, 99-169.

Hull, C. L. Aptitude Testing. Yonkers, N. Y.: World Book, 1928.

Hunter, J. E. & Schmidt, F. L. Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman & M. D. Dunnette (Eds.), Human performance and productivity: Volume 1. Human capability assessment. Hillsdale, N.J.: Erlbaum, 1982.

Hunter, J. E. & Schmidt, F. L. Quantifying the effects of psychological interventions on employee job performance and work-force productivity. American Psychologist, 1983, 78, 473-478.

Kelly, T. L. Statistical method. New York: Macmillan, 1923.

Schmidt, F. L., Hunter, J. E., McKenzie, R. & Muldrow, T. The impact of valid selection procedures on workforce productivity. Journal of Applied Psychology, 1979, 64, 609-626.

Taylor, H. L. & Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 23, 565-578.

Footnote

A preliminary version of this research was presented at the convention of the American Psychological Association, Los Angeles, California, in August 1983. Lawrence M. Hanser and Donald F. Haggard provided many helpful suggestions.

Table 1

Estimates of SD$ and Examples of Utility

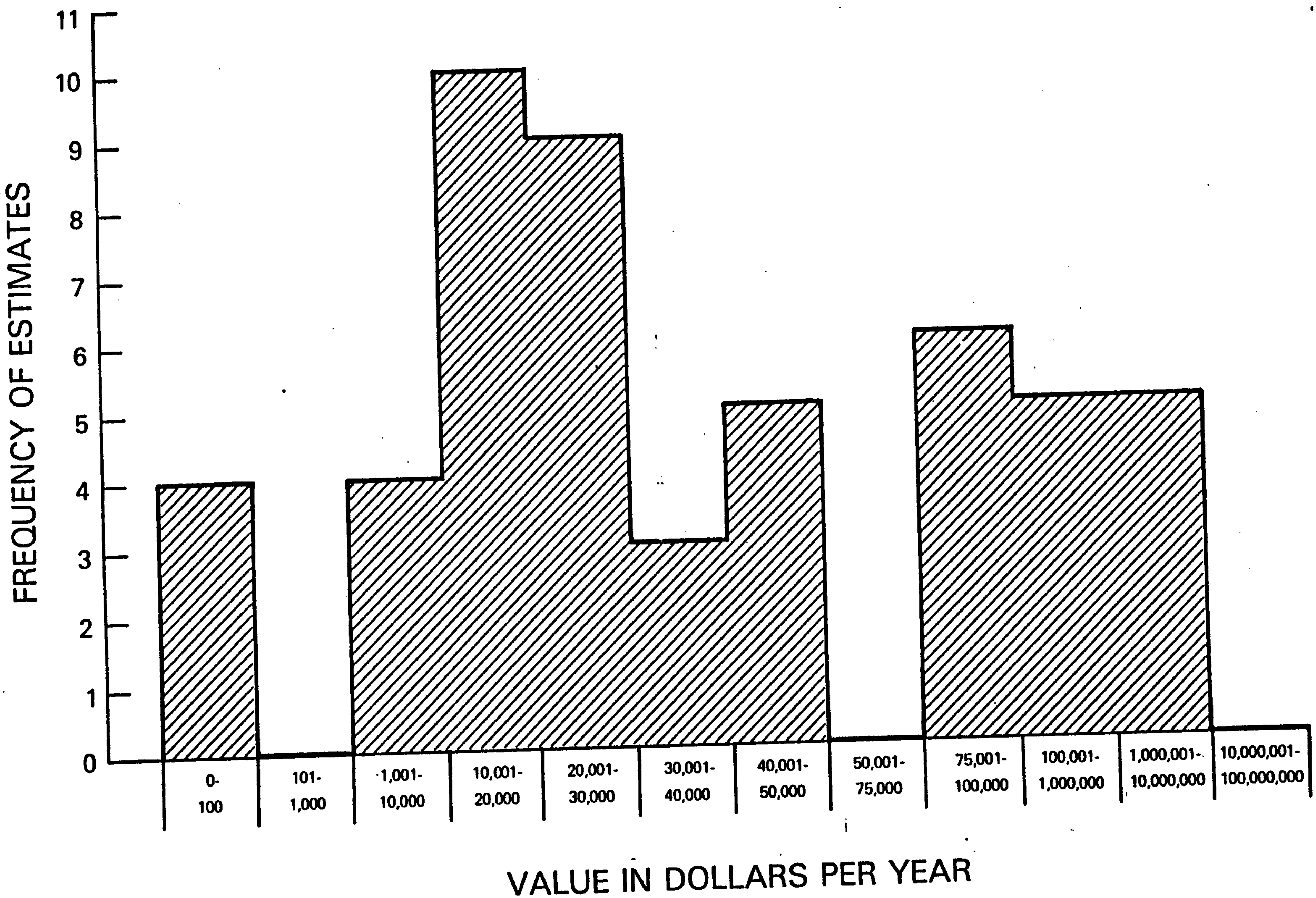
	SD$	U$ or utility per tank (Ns = 1)	U$ or utility per system (Ns = 2500)
SD$ Estimation Technique			
Difference between medians	$20,000	$ 4,800	$12,000,000
Median of differences	$15,000	$ 3,600	$ 9,000,000
Superior Equivalents Technique	$21,000	$ 5,040	$12,600,000
System Effectiveness Technique	$60,000	$14,400	$36,000,000

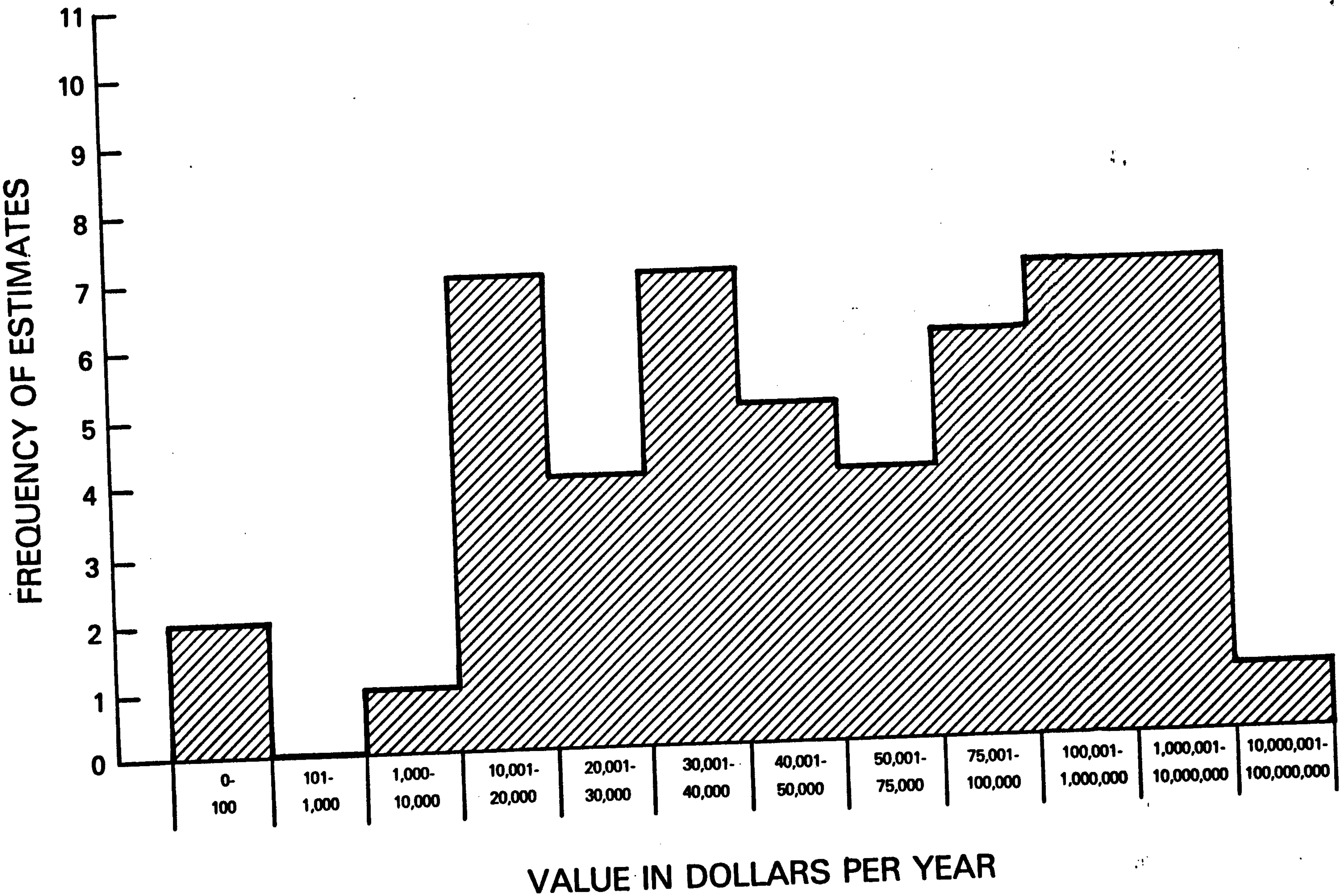
Figure Captions

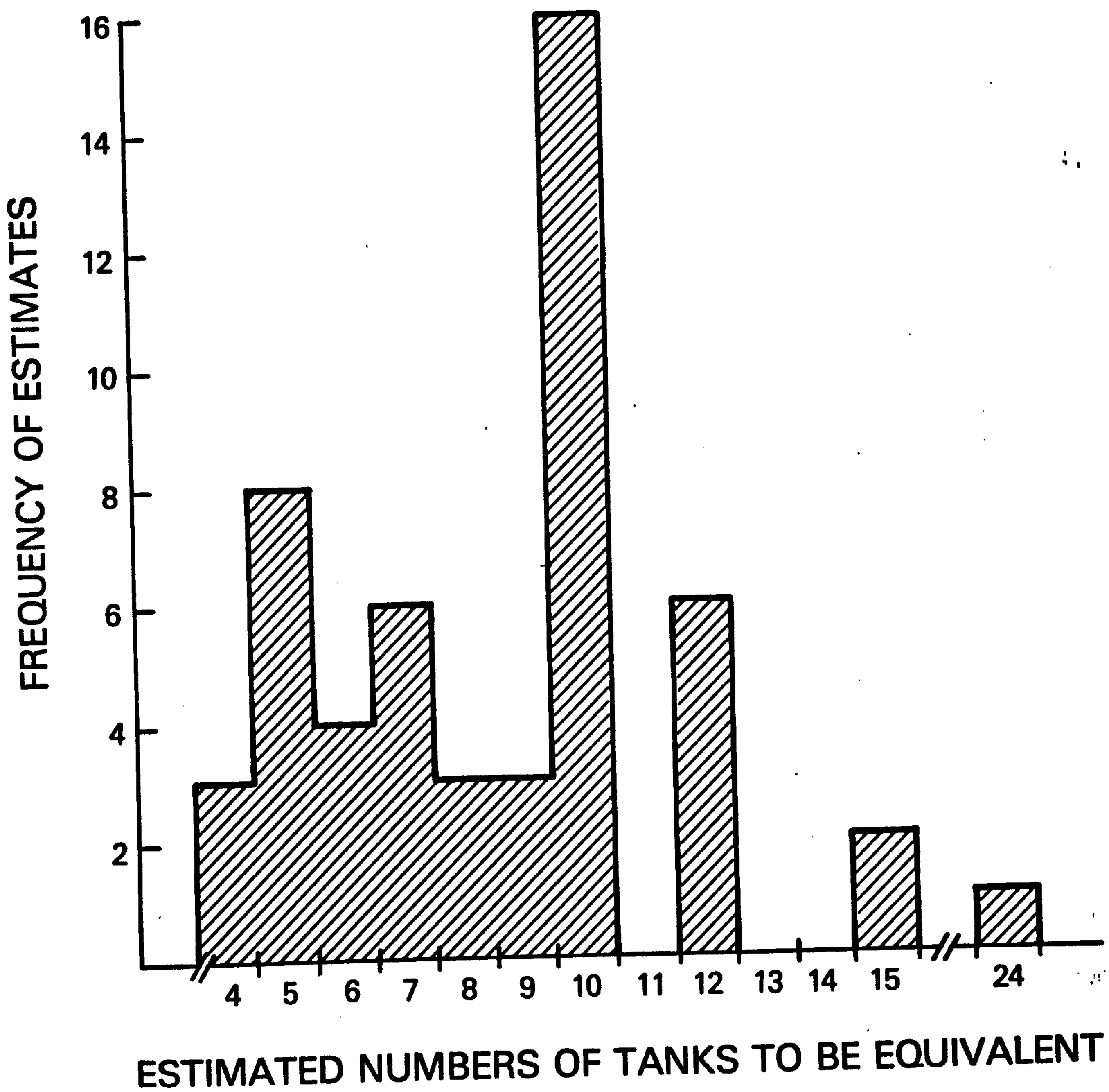
Figure 1. Estimation of value in dollars per year of an average tank commander.

Figure 2. Estimation of value in dollars per year of a superior tank commander.

Figure 3. Estimation of number of tanks with superior tank commanders equaling 17 tanks with average tank commanders.


FREQUENCY OF ESTIMATES
11
10
9
8
7
6
5
4
3
2
1
0
0-100
101-1,000
1,001-10,000
10,001-20,000
20,001-30,000
30,001-40,000
40,001-50,000
50,001-75,000
75,001-100,000
100,001-1,000,000
1,000,001-10,000,000
10,000,001-100,000,000
VALUE IN DOLLARS PER YEAR


FREQUENCY OF ESTIMATES
0
1
2
3
4
5
6
7
8
9
10
11
0-100
101-1,000
1,000-10,000
10,001-20,000
20,001-30,000
30,001-40,000
40,001-50,000
50,001-75,000
75,001-100,000
100,001-1,000,000
1,000,001-10,000,000
10,000,001-100,000,000
VALUE IN DOLLARS PER YEAR


FREQUENCY OF ESTIMATES
16
14
12
10
8
6
4
2
4
5
6
7
8
9
10
11
12
13
14
15
24
ESTIMATED NUMBERS OF TANKS TO BE EQUIVALENT

Working Paper

RS-WP-84-03

OFFICER SELECTION BATTERY FORMS 3 AND 4 MANUAL FOR ADMINISTRATION AND SCORING

Dorothy S. Edwards and John J. Claudy
AMERICAN INSTITUTES FOR RESEARCH

M.A. Fischl and Michael G. Rumsey
SELECTION AND CLASSIFICATION TECHNICAL AREA

October 1983

Reviewed by
Hilda Wing
Chief
Predictor Development Team

Francis F. Medland

Approved by
for Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

Selection and Classification Working Paper

OFFICER SELECTION BATTERY FORMS 3 AND 4

MANUAL FOR ADMINISTRATION AND SCORING

Dorothy S. Edwards and John J. Claudy
American Institutes for Research

M. A. Fischl and Michael G. Rumsey
U. S. Army Research Institute

October 1983

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Commercial: 202/274-8275
Autovon: 8/284-8275

I. GENERAL INSTRUCTIONS

1. Testing Method

The Officer Selection Battery, Forms 3 and 4 (OSB-3, DA and OSB-4, DA), is to be administered in accordance with the principles of test administration set forth in Chapter III, "Test Administration," in AR 611-5, Army Personnel Tests. In addition, the specific directions given in this manual for the OSB are to be followed without deviation. No omissions or changes in the wording of these directions are permitted. Each testing session is to be conducted by one qualified test administrator, and for groups greater than 25, one trained proctor for each additional 25 examinees.

Each form of the Officer Selection Battery consists of two separate test booklets, Part I and Part II. Each examinee is to be tested with Part I and Part II <u>of the same form</u>. The test forms are color coordinated; the part numbers are on the cover.

2. Order of Administration

Part I should be administered first, followed by Part II. Part I takes less time, so this sequence allows extra time for general introductory remarks that do not have to be repeated for the second part.

3. Time Required

About 50 minutes are required for administering each part of the Officer Selection Battery. Approximately 10 minutes are required for distributing test material, reading directions, and answering questions. The test is not timed, but examinees should be urged to work at a steady pace so they will finish during a class period.

The two test booklets will probably be administered in two separate class periods, though it is completely permissible to administer the two parts, one immediately after the other in a two-hour time period. If this is done, have the examinees stand up and walk around or stretch between the end of Part I and the beginning of Part II.

4. Materials Required for Each Test Period

A. For each examinee

1) One test booklet
2) One answer sheet
3) Two sharpened No. 2 pencils with erasers
4) Scratch paper (one sheet should be sufficient)

B. For the examiner

1) A copy of this manual
2) A test booklet
3) An answer sheet
4) Chalk and chalkboard or large paper for demonstrating answer marking, and for posting time remaining
5) Watch or clock
6) Extra pencils, extra scratch paper

5. <u>Test Security</u>

The test administratror will have the responsibility for security of the tests and answer sheets during the examination. Only those persons who will be involved in test administration should have access to these materials. When not in use, test material must be kept under lock and under control of a designated Test Control Officer.

6. <u>Advance Planning</u>

Advance planning will save time and make the test administration go smoothly. Read this manual before the test periods in order to improve the quality of test administration.

A. Be in the room when the examinees arrive. To minimize the opportunity for collaboration, arrange to separate examinees by seating them with empty seats between them, if possible.

B. Decide upon the quickest and most efficient way of distributing and collecting tests and answer sheets in the room where you will give the tests.

C. Plan how you will answer questions. In general, questions should be answered by rereading the test directions or instructing the examinee to reread them. <u>Do not amplify or interpret the directions.</u> A response to

further questioning about directions might be "Do the best you can, I am not permitted to tell you any more than what the printed directions say."

D. Plan to post the "Minutes Remaining" every 10 minutes on the chalkboard.

E. Motivating the examinees is very important. Some remarks have been prepared to be read to them, but if you also indicate by your attitude that <u>you</u> believe the test is important, it will help even more.

II. ADMINISTERING THE TESTS

1. Distribution of Testing Materials

Immediately prior to the testing session, place one set of testing materials, including one test booklet, one answer sheet, one sheet of scratch paper, and two pencils on each examinee's desk, face down.

2. Introductory Remarks by Test Administrator

Each testing session will be introduced by a set of prescribed remarks. All prescribed remarks are identified in this manual by vertical lines in the margins, and these passages MUST BE READ EXACTLY AS WRITTEN.

You will begin each test session by saying:

> You are about to take a test used by the Army to help select officer candidates and Advanced ROTC cadets. You will complete two test booklets, either both today or during two testing periods. These booklets will cover a number of different subject areas.
>
> One of the things the Army wants to know is what the best talents and skills of officer candidates are, so the tests cover many topics. Some questions will be easy and some will be hard. Don't be discouraged by the hard ones. Just work as quickly and accurately as you can.

Now pick up your answer sheet and turn it sideways, like this, so that the words LAST NAME are at the upper right. In the boxes at the top of Block 1, print your last name, leave a box blank, then your first name. Leave another box blank, and then print your middle initial. If your full name does not fit, just print as much as will fit. Use only a Number 2 pencil like the one you have been given.

Pause while examinees do this. Then say:

Now code each letter of your name by making a heavy black mark in the appropriate lettered space in the column below each letter of your name. If you have any questions, raise your hand.

Pause while examinees do this. Then say:

Now look at the lower left corner of the cover of your test booklet. There is a DA Form Number printed there. Code this number in Block 2 under DA Form on your answer sheet.

Pause while examinees do this. Then say:

Block 3 on your answer sheet, headed SSN, is for entering your Social Security Number. Enter your Social Security Number in the boxes at the top of Block 3. Omit hyphens, and record only the nine digits of the number. Then code each digit, including zeroes, by making a heavy mark in the appropriate numbered space in each column. Do this now.

Pause while examinees fill in this information. Then say:

Now open your test booklet and read the directions on the inside of the front cover while I read them aloud.

Read the directions printed inside the front cover of the test booklet. Then say:

Do you have any questions about what you are to do?

Answer any questions, but do not let the examinees delay the start of testing with irrelevant questions. Then say:

> If you need a new pencil or scratch paper during the test, raise your hand. You will have enough time to answer all questions, if you work steadily and don't spend too much time on any one question. Work quickly and accurately. Try to answer as many questions as you can. When in doubt, make your best guess. Ready? Turn to Page 1 and begin.

3. <u>Monitoring the Test Administration</u>

As examinees begin working, walk around the room and make sure that they have completed the top part of the answer sheet. Check especially that the proper test booklet number has been blackened and that the social security number has been recorded.

At about 10 minute intervals, write the minutes remaining on the chalkboard; e.g., "30 minutes left," 20, etc. The purpose in writing the minutes remaining is to aid examinees in keeping track of how the time of the class period is elapsing. It is expected that everyone will finish in the time allowed, but exceptionally slow examinees may not finish without prodding.

Any examinees who finish early should be instructed to re-check their work. After they have, the examiner at his/her option may allow them to turn in their materials and leave, but they must be prevented from creating a disturbance of those who a still working.

As examinees complete the tests, check their answer sheets t be sure the booklet number, Social Security Number, and other information at the top have been filled in. Collect their answer sheets and test booklets, placing them in <u>separate stacks</u>. Colle all scratch paper (to assure that none of the questions has been copied).

4. <u>After the Testing</u>

After testing has been completed and examinees dismissed, e test booklet should be scanned for stray marks or marked answers If the marks can be erased without any clue to answers remainin do so; otherwise destroy the booklet in accordance with procedu which will retain test security (burning or shredding).

All scratch paper should be destroyed.

III. SCORING THE OSB

1. Hand Scoring

Scan the answer sheet. If more than one answer to any question has been marked, draw a red line through all answer spaces to that question.

Place the scoring stencil that corresponds to the OSB Form number and Part number of the answer sheet on top of the answer sheet. Make certain that the stencil is properly aligned.

Count the number of black marks that show through the holes in the stencil. Do not count any marks that have a red line through them (remember, these are questions to which the examinee marked more than one answer, so none of them is counted as correct).

The total score--total number of right answers--should be written on the answer sheet in the place provided by the cut-out on the stencil. The scorer or another person should recount the black marks to make sure the total is correct.

2. Converting to Army Standard Scores

Use the appropriate conversion table to convert the examinee's raw score (total right) to an equivalent Army Standard Score. Be certain that you are using the proper table; it is the same color as the test booklets for the form used, and has the form number on it. If you have tested high school students, there is a separate conversion table for use with their scores.

Find the Raw Score in the left column, and read the corresponding Army Standard Score in the adjacent column. Record the Army Standard Score on the answer sheet, next to the raw score.

The Army Standard Score becomes the official Officer Selection Battery score, which is recorded in any official records or notices required by the Department of the Army, Department of Military Science or ROTC.

3. Machine Scoring

Selection and Classification Technical Area

Working Paper RS-WP-88-04

Preliminary Evaluation of Two Test-Based Strategies for Identifying Soldiers for the Excellence in Armor Program

ILENE F. GAST

February 1988

Approved for Distribution to Sponsor or Proponent.

REVIEWED BY:

Elizabeth P. Smith

ELIZABETH P. SMITH
Research Psychologist

APPROVED BY:

Lawrence M. Hanser

LAWRENCE M. HANSER
Chief, Selection and
Classification Technical Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Acknowledgements

The author would like to thank Dr. Elizabeth P. Smith for her help with the data analyses her careful reviews of this paper. Dr. Clinton B. Walker and Dr. Lawrence M. Hanser are also thanked for their helpful comments on earlier drafts.

EXECUTIVE SUMMARY

This paper describes the potential benefit of using information from Project A tests in selecting soldiers for the Excellence in Armor Program (EIA). To evaluate the impact of the new tests, we compared end-of-training (EOT) measures for EIA soldiers in the Project A database with those of non-EIA soldiers (19E and 19K) and with samples formed from the same pool by applying two test-based decision rules. Finally, we examined the effects on EOT measures of applying the decision rules to the EIA soldiers.

Two different test-based rules were generated. Each was devised to designate approximately the top 30 percent of the pool of soldiers earmarked for 19 CMF. One strategy emphasized the Assessment of Background and Life Experiences (ABLE), a measure of leadership potential (ABLE Emphasis); another equally emphasized ABLE and a set of spatial and psychomotor tests (Equal Emphasis). One important difference between the existing and test-based strategies is the point at which information becomes available to decision makers. The existing strategy relies in information collected after soldiers have begun OSUT. The new strategy generates information before training begins.

The new test-based rules were evaluated in a sample of 2,372 soldiers in MOS 19E and 19K. All soldiers had graduated from One Station Unit Training (OSUT), 162 of these in the EIA program. In addition, these soldiers had been tested on Project A predictors as receptees and had completed EOT measures. The EOT measures included peer and supervisory ratings and scores on an MOS-specific EOT knowledge test. We first applied the new rules to the total sample of 2,372 soldiers and then the smaller subsample of 162 EIA graduates. Our purpose was to compare the two new strategies with each other and with the existing strategy.

We found that the existing strategy for tracking trainees is working well. It produces EIA graduates who are seen as superior to Normal Track (NT) soldiers by their superiors and peers and who outscore their NT counterparts on EOT tests.

The soldiers designated by the new strategies scored similarly to the EIA graduates on the EOT tests and about halfway between the ET and NT sample on the peer and supervisory ratings. The Equal Emphasis rule proved superior to ABLE Emphasis rule due, in part, to the addition of the predictive ability of the spatial and psychomotor tests. In the full sample, both decision rules produced similar peer and supervisory ratings but the Equal Emphasis rule produced better performance on the EOT tests. In addition, the Equal Emphasis rule created a larger pool of soldiers than did the ABLE Emphasis rule.

We also applied the new test-based strategies to the EIA sample to determine how closely they replicated existing selection procedures. The new test-based strategies selected only 40-50 percent of those selected by existing strategies. However, those EIA graduates who were selected by the test-based strategies were superior to the remaining soldiers in all

respects. Thus, the new tests selected the "cream of the crop" within EIA.

Based on our evaluation, we conclude that the existing information sources and new test-based information are complementary and appear to work well together to select the best EIA candidates. We also found that a test-based strategy giving equal weight to the ABLE and a composite of the spatial and psychomotor tests was superior to one emphasizing the ABLE alone. As in past research, we expect that the spatial/psychomotor component of the Equal Emphasis strategy will yield improvement in the "can do" hands-on gunner performance (such as that measured by the UCOFT), and that the ABLE component will predict the "will-do" components of performance.

Introduction

The purpose of this paper is to provide information to the U.S. Army Armor School to support their planning for the Gunner Selection and Sustainment Program. The Excellence in Armor Program (EIA) calls for the identification of outstanding soldiers in the Armor and Cavalry field. It was created to provide incentives to encourage these soldiers' continued superior performance and improve their retention rate. Soldiers are identified for the Excellence Track (ET) in their seventh week of One Station Unit Training (OSUT) for early advancement and specialized developmental experiences. This paper compares attributes and performance of soldiers who are in the Excellence in Armor Program to those in the Normal Track (NT). In addition, it evaluates the effect of two potential rules for making decisions based on the Project A tests.

Method

Subjects

Subjects were 2,372 soldiers in the 19 CMF (578 in MOS 19E and 1,794 in MOS 19K) who entered the Reception Battalion at Ft. Knox during the period from August 1986 to November 1987 and subsequently completed the full Project A predictor battery. Of these soldiers, 162 have since completed OSUT in the Excellence Track and 2,210 have completed the Normal Track.

Measures of Soldiers' Attributes

Cognitive Ability. Before entering military service, all soldiers completed the Armed Services Vocational Aptitude Battery (ASVAB). The Armed Forces Qualification Test (AFQT), a composite of four of the ten ASVAB subtests, served as the measure of cognitive ability.

Psychomotor Skills. Two computerized psychomotor tests were included from the Project A Battery: One-handed Tracking and Two-handed Tracking. The former measures steadiness and precision; the latter measures coordination and dexterity.

Spatial Ability. Two of the Project A paper-and-pencil tests of spatial ability were included. The Maze Test assesses the ability to scan a field. The Orientation Test measures the ability to maintain one's perspective or bearing with respect to some object when it and its component parts have been rotated.

Spatial and psychomotor test scores were formed into a single composite score.

Leadership Potential. Leadership Potential was assessed using four scales from the Assessment of Background and Life Experiences (ABLE), one of the noncognitive tests included in the Project A predictor battery.

The four ABLE scales were: Dependability, Adjustment, Work Orientation and Social Dominance. A composite was formed by summing the scores obtained on the 68 items included in these scales.

Performance Measures

As part of ongoing Project A data collection, soldiers' performance was assessed at the end of OSUT. This subset of Project A criterion measures was used in the present analyses. These measures are described below.

Ratings. Supervisors and peers rated soldiers on a series of behaviorally anchored rating scales. Ratings were made on a seven-point scale. Dimensions included: Technical Skill, Effort, Following Regulations and Orders, Military Appearance, Physical Fitness, Self-Control, and Leadership Potential.

End-of-Training (EOT) Knowledge Tests. At the end of training, soldiers completed an MOS-specific knowledge test addressing the material covered in training.

Development of Decision Rules

New test-based decision rules were developed for evaluating the relation of scores on the Project A tests to selection into EIA and to measures of performance. In developing these rules, we considered the two kinds of abilities assessed by the Project A predictors: leadership potential (as measured by the ABLE) and spatial/psychomotor abilities (as measured by the two paper-and-pencil tests of spatial ability and two computerized tests of psychomotor skills). We formed our decision rules by examining flow rates into the 19 CMF. Norm tables for these decision rules were based on the scores of over 20,000 soldiers in combat MOS (11B, 13B, 16S, 19E and 19K) who took these tests in Project A during the period from August 1986 to November 1987.

Two different strategies were investigated, each with the goal of designating the top 30 percent of receptees. We selected 30 percent as a target because doing so would create a large enough pool to accommodate practical concerns such as the variable flow of soldiers into the Reception Battalion, and the need to identify a large enough pool of highly qualified soldiers to permit selection of top quality soldiers for other programs as well. The two strategies differed in the amount of influence afforded to each of the two kinds of measures.

One decision strategy, (Equal Emphasis) gave equal weight to the ABLE and the spatial/psychomotor composite. To be included in the pool, soldiers had to have scores on both sets of tests comparable to the top 55 percent of the 20,000 soldiers in the comparison group (i.e., above the 45th percentile on each kind of test). In effect, this rule assigns equal importance to gunnery aptitude and to leadership potential.

A second strategy (ABLE Emphasis) afforded more weight to the ABLE. This strategy formed a pool of soldiers who scored well on the ABLE, while screening out those with poor spatial and psychomotor skills. To be included, soldiers had to score better on the ABLE than 67 percent of the soldiers in the comparison group (i.e., above the 67th percentile). In addition, this decision rule excluded soldiers who scored below the 10th percentile on a composite formed from scores on the spatial and psychomotor tests. Thus, this rule stresses leadership potential but excludes examinees who have very low aptitude for gunnery.

Analyses

One purpose of the analyses was to compare the pool of soldiers who were selected for EIA using existing strategies with pools resulting from each of the two decision rules described above. For this comparison, subgroups were created by applying each decision rule to the total sample ($N = 2,372$). The Equal Emphasis decision rule produced a pool of 746 soldiers; the ABLE Emphasis decision rule produced a pool of 592. A third subgroup consisted of EIA graduates ($N = 162$). As a basis of comparison, those soldiers who did not graduate from EIA are also included ($N = 2,210$). Means on predictor and criterion variables were then computed.

A second purpose of the analyses was to compare the two new decision rules to the existing one. This was done by applying the decision rule to the sample of EIA graduates to see how many would be included by each rule.

Results

Table 1 shows means on predictor and criterion variables for the four different subgroups. The existing decision strategy and the two test-based strategies select soldiers who are clearly superior in all respects to those who were not EIA graduates. Some differences emerged among the pools resulting from the three selection strategies. As would be expected, the Equal Emphasis decision rule produced the highest scores on the spatial/psychomotor composite and the ABLE Emphasis rule produced the highest ABLE scores. Compared with unselected NT soldiers, OSUT graduates meeting either the Equal Emphasis or ABLE Emphasis decision rule scored consistently higher on the EOT ratings and tests. Compared with Equal Emphasis, ABLE Emphasis produces slightly higher means on the ratings. Conversely, Equal Emphasis produces a slightly higher score on the Training Knowledge Test. The existing selection procedure produces performance ratings that are consistently higher than those produced by the test-based strategies; means on the Training Knowledge test are comparable to those produced by the Equal Emphasis rule.

The next step in the analyses was to apply each of the two new test-based decision rules to the EIA sample ($N = 162$). The two test-based decision rules do not reproduce existing decision strategies. The Equal

Emphasis rule would result in the selection of 77 EIA graduates (about 50% of the total); ABLE Emphasis would select 66 (42%) of the EIA graduates.

Table 2 compares EIA graduates who would have been selected using the new decision rules to those who would not have been selected and to the full subsample of EIA graduates. Both decision rules select soldiers who tend to score better than the average EIA graduate on the predictors and criteria. In all respects, those who would have been selected using the Equal Emphasis rule are superior to those who would not have been selected. The most notable differences are in the three predictor scores, the peer and supervisory ratings of technical skill and performance on the Training Knowledge Test. The ABLE Emphasis rule produces similar but less striking results.

Discussion

The present investigation compared existing procedures for tracking trainees with two new test-based decision strategies. The first strategy, Equal Emphasis includes soldiers whose scores are comparable to those in the top 55 percent of the comparison group on both the ABLE and a composite of spatial and psychomotor tests. The second strategy, ABLE Emphasis, selects soldiers who are comparable to the top third of the comparison group on ABLE and screens out those comparable to the bottom tenth on the spatial and psychomotor composite.

The data suggest that the existing strategy for tracking trainees is successfully producing a group of EIA graduates who are seen as superior by their supervisors and peers and who score well on EOT tests. However, the data also suggest that much can be gained from incorporating the Project A Tests into existing strategies.

The existing strategy and the new test-based rules use different sources of information in designating soldiers. Therefore, it is not surprising that the existing and test-based strategies would produce visibly different results. The new test-based rules rely solely on information available before OSUT, whereas the existing strategy takes advantage of seven additional weeks of performance-based information from training context.

In some ways, our comparison between existing and test-based strategies is stacked in favor of the existing strategy. The peer and supervisory ratings may have been positively influenced by raters' knowledge that certain individuals had been targeted for EIA. Further, ET soldiers received developmental experiences during training that were not made available to soldiers in the NT. In addition, the EIA sample contains only those who successfully completed the program. A fair percent of those who enter EIA during OSUT drop out before training ends. If we had been able to obtain predictor test scores and EOT measures for all soldiers who had been in the EIA program (including dropouts), their scores on the peer and supervisory measures may have been more similar to those of soldiers selected with one of our new decision rules.

Our application of the new test-based decision rules to the 162 EIA graduates served, in part, as a control for the effects of criterion contamination. All soldiers in this group had been exposed to similar in-training influences. Within the group of EIA graduates, we compared predictor and criterion scores of those who would have been selected by each rule to those who would not have been selected. When the new decision rules are applied to the EIA graduate sample, the best EIA graduates (the "cream of the crop") are selected. (See Table 2.) The means on Table 2 suggest that an Equal Emphasis strategy appears to be superior to ABLE Emphasis. As an added advantage, Equal Emphasis creates a larger pool of eligible soldiers.

In conclusion, our intent is not to work against the present selection system, but to point out that the test scores, when used in conjunction with existing procedures, can produce an even better pool of soldiers for the Excellence Track. We expect that a selection process which emphasizes spatial and psychomotor abilities equally with leadership potential, like the Equal Emphasis rule, will provide additional advantages later in soldiers' careers. Past research (Smith & Graham, 1987) suggests that this strategy will yield improvement in the "can do" hands-on gunner performance, such as that measured by the UCOFT, because of the additional emphasis it gives to the spatial and psychomotor tests. Furthermore, the ABLE predicts the "will-do" components of performance (Hough, Gast, White & McCloy, 1986). Thus, we can expect soldiers who have higher ABLE scores to be more willing to perform in training and on the job.

Table 1: Comparison of Means on Predictors and Criteria: New Decision Rules vs. Existing Strategy

Variable	Non-EIA N = 2,210	Equal Emphasis N = 753	ABLE Emphasis N = 592	EIA N = 162
Predictors				
AFQT	59	67	65	64
ABLE[a]	44	74	86	57
Spatial/Psychomotor[b]	51	78	65	65
Criteria				
Peer Ratings				
Technical Skill	4.35	4.76	4.72	5.23
Effort	4.03	4.28	4.27	4.41
Following Regs./Orders	4.25	4.53	4.47	4.83
Military Appearance	4.53	4.73	4.75	5.01
Physical Fitness	4.51	4.71	4.74	5.18
Self-Control	4.34	4.49	4.43	4.40
Leadership Potential	3.79	4.20	4.46	4.65
Supervisory Ratings				
Technical Skill	4.84	5.15	5.18	5.69
Effort	4.72	5.02	5.05	5.53
Following Regs/Orders	4.81	5.08	5.13	5.62
Military Appearance	4.72	4.95	5.09	5.44
Physical Fitness	4.66	4.87	5.02	5.50
Self-Control	4.83	5.07	5.15	5.57
Leadership Potential	4.41	4.78	4.92	5.51
Training Knowledge Test	104.35	112.62	110.90	112.51

[a]Means on the ABLE and Spatial/Psychomotor Composite are expressed in terms of percentiles.

[b]The Spatial/Psychomotor score is a weighted composite of the scores from the Spatial and Psychomotor tests.

Table 2: Comparison of Means on Predictors and Criteria: Application of New Decision Rules to EIA Graduates

Variable	Total EIA Grad. N = 162	Equal Emphasis Included N = 77	Equal Emphasis Excluded N = 80	ABLE Emphasis Included N = 66	ABLE Emphasis Excluded N = 91
Predictors					
AFQT	64	70	59	68	62
ABLE[a]	57	76	39	85	35
Spa/Psyc[b]	65	79	48	66	62
Criteria					
Peer Ratings					
Technical Skill	5.23	5.43	5.05	5.35	5.14
Effort	4.41	4.49	4.34	4.60	4.27
Follow Regs./Orders	4.83	4.93	4.75	4.83	4.83
Military Appearance	5.01	5.05	4.97	5.00	5.01
Physical Fitness	5.18	5.13	5.22	5.12	5.22
Self-Control	4.40	4.44	4.37	4.33	4.45
Leadership Pot.	4.65	4.72	4.59	4.71	4.61
Supervisory Ratings					
Technical Skill	5.69	5.89	5.52	6.02	5.46
Effort	5.53	5.77	5.32	5.75	5.37
Follow Regs/Orders	5.62	5.72	5.54	5.88	5.44
Military Appearance	5.44	5.64	5.27	5.68	5.27
Physical Fitness	5.50	5.59	5.43	5.63	5.41
Self-Control	5.57	5.67	5.48	5.67	5.50
Leadership Pot.	5.51	5.64	5.40	5.67	5.40
Trng. Knowledge Test	112.51	117.45	108.33	115.64	110.27

[a]Means on the ABLE and Spatial/Psychomotor Composite are expressed in terms of percentiles.

[b]The Spatial/Psychomotor score is a weighted composite of the scores from the Spatial and Psychomotor tests.

References

Hough, L. M., Gast, I. F., White, L. A., & McCloy, R. (1986, August). The relation of leadership and individual differences to job performance. Paper presented at the Meeting of the American Psychological Association, Washington, D. C.

Smith, E. P., & Graham, S. E. (1987). Validation of psychomotor and perceptual predictors of Armor Officer M-1 gunnery performance. (ARI Technical Report No.). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Selection and Classification Technical Area Working Paper RS-WP-88-10

COMPUTING ARMY APTITUDE AREA COMPOSITE SCORES (1980 METRIC) FOR ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB), FORMS 6 AND 7

Frances C. Grafton

October 1988

Approved for Distribution to Sponsor or Proponent.

REVIEWED BY:

Mark Y. Czarnolewski
Research Psychologist

APPROVED BY:

Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Computing Army Aptitude Area Composite Scores (1980 metric) for the Armed Services Vocational Aptitude Battery (ASVAB), Forms 6 and 7

INTRODUCTION

In October 1984 the reference population for all military aptitude tests was changed from the 1944 mobilization population to the 1980 youth population. At that time four new versions of the operational Armed Service Vocational Battery (ASVAB 11/12/13/14) and a single form of the in-service test battery, the Armed Forces Classification Test (AFCT-9), were issued with conversion tables appropriate for the new reference population.

Since only one form of the in-service test is currently in use, the Army has no viable recourse in the case of potential compromise of the AFCT.

The last form of the AFCT, prior to the change in reference populations was, a renormed version of ASVAB 7B. All forms of ASVAB 6/7 used conversion tables based on the 1944 mobilization population as the reference population. The purpose of this research is to produce new conversion tables, appropriate to 1980 youth reference population, to be used with AFCT-7B.

BACKGROUND

With the introduction of ASVAB 11/12/13/14 the reference population for military aptitude tests became the 1980 youth population of all 17-23 year old youth (both male and female). In order to compare scores received prior to October 1984 to those received afterward, it was necessary to develop a method to put scores on a common reference population metric (preferably the 1980 metric). Since the test administered to the youth population, ASVAB 8A, had been previously normed to the 1944 mobilization population (Maier & Grafton, 1980), conversion tables between standard scores in the two metrics could be directly computed from the youth population data itself. Tables converting AFQT percentile scores as well as service-specific aptitude area composite standard scores from the 1944 metric to the 1980 metric were published just prior to the implementation of the 1980 metric in the operational ASVAB (Mitchell and Hanser, 1984; Sims, et. al., 1984).

PROBLEM

Generally only one form of the Armed Forces Classification Test is available for in-service testing. The ASVAB form to be used is selected by an inter-service committee. The choice is made among forms that are no longer in operational use. The current AFCT is the only ASVAB form from the ASVAB 8/9/10

series that has not been designated for another specific purpose. If an interim replacement for the current AFCT is to be used, it must be selected from the ASVAB 6/7 series. Conversion tables for the recalibrated versions of ASVAB 5/6/7 were published in July 1980 (Department of Defense, 1980). These tables were calibrated to 1944 mobilization population. In order to use any form of ASVAB 6/7 as an alternate form of the AFCT, new conversion tables need to be developed so that scores will be computed in the 1980 metric.

METHOD

The computational formulas for the 10 Army Aptitude Area Composites are from the revised conversion tables for ASVAB 6/7/6E/7E (Department of Defense, 1980). Table A1 of the appendix lists the formulas and the composites. Raw composite scores for ASVAB 6/7 are sums of subtest raw scores. Once the sum of subtest raw scores is computed for each aptitude area score, the corresponding aptitude area composite standard score is obtained by a table look up procedure.

Each of the composite standard scores in Tables 1 and 2 of DoD 1304.12W1 (Revised July 1980), was converted to the equivalent 1980 aptitude area composite score by using Table A2 in Sims, et. al. (1980). Tables A2 and A3 of the Appendix provide direct conversions for the ASVAB 6/7 raw composite scores to 1980-metric aptitude area standard scores. The original 1944-metric are not provided in Tables A2 and A3 in order to prevent errors in conversion should the tables be used in actual scoring.

The Armed Forces Qualification Test (AFQT) score is not recomputed when the AFCT is taken for record. For this reason, tables for converting AFQT scores from the 1944 metric to the 1980 metric are not provided. These tables are available, however, in both Mitchell and Hanser (1984) and in Sims, et. al. (1984).

DISCUSSION AND CONCLUSIONS

If the Armed Forces Classification Test (AFCT) needs to be replaced due to compromise of the current version, the only possible candidates are from the ASVAB 6/7 series. AFCT form 7B appears to be a likely replacement. If AFCT-7B or any other form from the ASVAB 6/7 series is to be used operationally, the scores should be reported in the 1980 metric. Tables A1 - A3 of the appendix provide a procedure for accomplishing this goal and should be issued to the field as appropriate.

References

Department of Defense (1980). Conversion Tables for ASVAB 6/7/6E/7E. (DoD 1304.12W1 (Revised July 1980). Washington, DC: Department of Defense.

Maier, M. H. & Grafton, F. C. (1980). Renorming ASVAB 6/7 at Armed Forces Examining and Entrance Stations. (Technical Memorandum 80-1) Washington, DC: Office of the Secretary of Defense.

Mitchell, K. J. & Hanser, L. M. (1984). The 1980 Youth Population Norms: Enlistment and Occupational Classification Standards in the Army. (Working Paper RS-WP-84-07), Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sims, W. H., Maier, M., Stoloff, P. H. Truss, A. R., & Hiatt, C. M. (1984). Converting Military Test Scores from the WWII Score Scale into the 1980 Score Scale. (CNA Memorandum 83-3076.10). Alexandria, VA: Center for Navel Analyses.

1 October 1988

TABLE A1

COMPUTATION FORMULAS[a]
ARMY APTITUDE AREA COMPOSITES
ARMED FORCES CLASSIFICATION TEST (JAN 82)

COMPOSITE	FORMULA
General Technical	GT=AR+WK
General Maintenance	GM=AR+GS+MC+AI
Electronics	EL=AR+EI+MC+SI+CE
Clerical	CL=AR+WK+AD+CA
Mechanical Maintenance	MM=MK+SI+EI+AI+CM
Surveillance/Communications	SC=AR+WK+MC+SP
Combat	CO=AR+SI+SP+AD+CC
Field Artillery	FA=AR+GI+MK+EI+CA
Operators/Food	OF=GI+AI+CA
Skilled Technical	ST=AR+MK+GS

[a] Composites are computed as sum of subtest raw scores.

TABLE A2
ARMY TABLES FOR COMPOSITES
ARMED FORCES CLASSIFICATION TEST, FORM 7B

	Composite Standard Score[a]					
Raw Score	GT	GM	EL	CL	MM	Raw Score
105-110			136		137	105-110
104			136		136	104
103			136		136	103
102			136		135	102
101			136		135	101
100			135	129	134	100
99			135	129	134	99
98			133	129	133	98
97			133	129	132	97
96			132	129	132	96
95			132	129	131	95
94			132	129	130	94
93			131	129	129	93
92			131	129	127	92
91			130	129	127	91
90			129	129	127	90
89			127	128	126	89
88			127	128	125	88
87			125	128	124	87
86			124	128	123	86
85			122	128	123	85
84			121	127	121	84
83			120	127	121	83
82			118	125	119	82
81			118	125	119	81
80			118	125	118	80
79			116	123	117	79
78		138	115	122	116	78
77		136	114	122	116	77
76		136	113	121	115	76
75		136	113	120	115	75
74		135	111	118	113	74
73		134	111	116	113	73
72		134	110	115	112	72
71		133	109	114	110	71
70		132	107	114	109	70
69		132	106	112	109	69
68		131	105	112	108	68
67		130	103	111	107	67
66		129	102	108	105	66

TABLE A2 (Continued)

Composite Standard Score

Raw Score	GT	GM	EL	CL	MM	Raw Score
65		127	101	108	104	65
64		124	100	108	104	64
63		123	99	107	102	63
62		122	98	106	101	62
61		121	97	105	100	61
60		119	96	103	99	60
59		118	96	102	98	59
58		118	93	102	96	58
57		116	92	99	95	57
56		115	91	98	95	57
55		115	90	97	93	55
54		114	89	95	92	54
53		114	88	93	91	53
52		112	88	90	89	52
51		112	87	89	88	51
50	128	111	85	89	88	50
49	127	109	84	86	87	49
48	125	108	84	85	86	48
47	124	107	83	84	84	47
46	123	105	81	82	84	46
45	121	103	80	81	82	45
44	118	102	80	79	81	44
43	116	101	79	77	79	43
42	116	99	77	74	78	42
41	115	98	76	72	77	41
40	113	97	75	70	76	40
39	112	96	74	68	74	39
38	111	94	73	67	73	38
37	109	92	72	64	72	37
36	109	91	71	62	71	36
35	108	89	71	61	70	35
34	103	87	70	59	68	34
33	102	87	68	58	67	33
32	101	85	67	58	66	32
31	99	83	67	56	64	31
30	97	81	64	52	63	30
29	95	80	63	50	62	29
28	93	79	63	49	59	28
27	89	77	63	48	57	27
26	[illegible]	76	62	48	57	26

TABLE A2 (Continued)

Raw Score	GT	GM	EL	CL	MM	Raw Score
	Composite Standard Score					
25	86	74	62	48	57	25
24	83	73	62	48	57	24
23	80	71	62	48	57	23
22	78	70	62	48	57	22
21	77	69	62	48	57	21
20	75	68	62	48	57	20
19	73	67	62	48	57	19
18	71	64	62	48	57	18
17	70	60	62	48	57	17
16	68	60	62	48	57	16
15	66	60	62	48	57	15
14	64	60	62	48	57	14
13	61	60	62	48	57	13
12	61	60	62	48	57	12
11	58	60	62	48	57	11
0-10	57	60	62	48	57	0-10

[a]All standard scores are in the 1980 metric.

1 October 1988

TABLE A3
ARMY TABLES FOR COMPOSITES
ARMED FORCES CLASSIFICATION TEST, FORM 7B

Composite Standard Score[a]

Raw Score	SC	CO	FA	OF	ST	Raw Score
110-117		137				110-117
109		137				109
108		136				108
107		135				107
106		135				106
105		135	136			105
104		133	136			104
103		133	136			103
102		133	136			102
101		133	136			101
100		133	136			100
99		133	136			99
98		132	136			98
97		132	136			97
96		130	134			96
95		129	134			95
94		128	133			94
93		127	133			93
92		126	132			92
91		125	131			91
90	132	133	130			90
89	132	122	130			89
88	132	121	129			88
87	132	119	128			87
86	131	118	128			86
85	131	117	127			85
84	129	116	125			84
83	128	115	124			83
82	127	114	124			82
81	126	113	123			81
80	125	112	123			80
79	124	111	120			79
78	123	110	119			78
77	122	109	118			77
76	120	106	116			76
75	119	104	115			75
74	118	103	114			74
73	118	102	113			73
72	117	100	111			72
71	116	99	110			71

TABLE A3 (Continued)

Raw Score	SC	CO	FA	OF	ST	Raw Score
	Composite Standard Score					
70	115	98	110			70
69	114	96	109			69
68	113	94	108			68
67	113	93	106			67
66	112	91	105			66
65	111	89	104			65
64	111	89	102			64
63	110	87	101			63
62	110	86	99			62
61	108	86	99			61
60	108	83	98		135	60
59	106	82	97		135	59
58	104	81	95		135	58
57	104	80	94		133	57
56	103	79	91		132	56
55	103	77	91	134	131	55
54	101	75	90	134	129	54
53	100	74	88	134	128	53
52	99	74	87	134	126	52
51	98	71	86	134	124	51
50	96	69	84	134	123	50
49	94	69	83	134	122	49
48	93	68	83	134	121	48
47	91	67	82	134	120	47
46	90	66	81	132	117	46
45	89	65	80	130	116	45
44	87	64	78	129	115	44
43	85	62	78	129	114	43
42	85	60	75	125	113	42
41	82	60	75	124	112	41
40	81	58	74	121	110	40
39	80	58	73	119	109	39
38	78	57	72	118	109	38
37	75	57	70	116	107	37
36	74	57	70	115	106	36
35	71	57	69	114	104	35
34	70	57	67	111	103	34
33	68	57	66	108	102	33
32	66	57	65	107	99	32
31	64	57	64	104	98	31

TABLE A3 (Continued)

	Composite Standard Score					
Raw Score	SC	CO	FA	OF	ST	Raw Score
30	63	57	63	101	97	30
29	61	57	62	99	94	29
28	59	57	59	96	92	28
27	56	57	58	93	90	27
26	53	57	58	89	89	26
25	50	57	58	88	86	25
24	49	57	58	86	85	24
23	48	57	58	84	83	23
22	48	57	58	81	81	22
21	48	57	58	77	79	21
20	48	57	58	75	78	20
19	48	57	58	71	74	19
18	48	57	58	68	71	18
17	48	57	58	66	69	17
16	48	57	58	63	67	16
15	48	57	58	61	65	15
14	48	57	58	58	63	14
13	48	57	58	55	62	13
12	48	57	58	53	59	12
11	48	57	58	53	58	11
0-10	48	57	58	53	50	0-10

[a]All standard scores are in the 1980 metric.

Selection and Classification Technical Area Working Paper RS-WP-87-10

RELATIONSHIP BETWEEN PROJECT A PSYCHOMOTOR AND SPATIAL TESTS AND TOW2 GUNNERY PERFORMANCE: A PRELIMINARY INVESTIGATION

Frances C. Grafton
Mark Y. Czarnolewski
Elizabeth P. Smith

January 1988

REVIEWED BY:
Michael G.Rumsey

APPROVED BY:
Lawrence M. Hanser
Chief, Selection and
Classification Tech Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

ACKNOWLEDGMENTS

We would like to express our appreciation to the following people for providing assistance with this project:

Army Research Institute Field Unit-Fort Benning
- Dr. Seward Smith, Chief
- Dr. Jean Dyer
- Dr. Georgann Lucariello

U.S. Army Infantry School, Fort Benning

Mr. James Callaway	DOES-USAIS
Mr. Jan Chervenak	DOES-USAIS
Mr. Elton Jones	DOTD_USAIS
CPT Eric Lynam	DOES-USAIS

29th Infantry Regiment

CPT Marty Russo	CoB, 2/29th INF (COR)
SGT John Medford	NCOIC, 2nd Battalion, B Company

Litton Computer Services
- Mr. Ron Mertere
- Mr. Woody Greene

American Institutes for Research
- Ms. Winnie Young

Relationship between Project A Psychomotor and Spatial Tests and TOW2 Gunnery Performance: A Preliminary Investigation

INTRODUCTION

Within the Army, matching the right person to the right job continues to be an important issue. Ongoing research efforts attempt to optimize person-to-job matches. One such major effort is the Army's Project A: Improving the Selection, Classification and Utilization of Army Enlisted Personnel. One of Project A's principal goals is to improve traditional methods and add new ones to enable measurement of a greater range of abilities required for successful soldier performance (Eaton, Goer, Harris, & Zook, 1984). This work has resulted in development of a computerized and pencil-and-paper test battery of temperament, cognitive, psychomotor, perceptual and spatial variables which are expected to predict performance in a diversity of Military Occupational Specialties (MOS). Early results indicate that the Project A battery has incremental validity in predicting enlisted performance over that obtained with the Armed Services Vocational Aptitude Battery (ASVAB) alone, the single current selection/classification tool.

As work on Project A has progressed, it has become apparent that there are special applications for which parts of the battery have considerable potential. Among the most apparent is the use of psychomotor and spatial tests for the selection of gunners. The tests have already been researched with M-1 tank gunners using the Unit Conduct of Fire Trainer (UCOFT) (Smith & Graham, 1987). The high correlation between the tests and performance ($\underline{R}$ = .76) suggests that similar results with other gunnery specialties are possible. The present effort extends research on the validity of the Project A tests for assigning basic trainees in Infantry into Advanced Individual Training (AIT) for MOS 11H (TOW gunner). The work was undertaken at the request of the Commanding General of the Training and Doctrine Command (TRADOC).

Background

The advent of anti-armor systems that fire wire-guided missiles for Infantry use brought new considerations to the issues of personnel selection and training (Cartner, Strasel, Evans, Heller, & Tierney, 1985). Both the heavy anti-tank TOW (Tube-launched, Optically-tracked, Wire-guided) and weapons of the "fire and forget" category such as the bazooka and M-1 tank require abilities of target acquisition and identification, prediction of movement direction, and judgment of when to fire. The TOW, however, also requires that the gunner maintain the missile on target until the missile hits, i.e., continuous sight alignment. The ground mount TOW2, the system of interest in this research, requires the gunner to manipulate knobs with both hands (one for the horizontal axis and one for the vertical axis) while visually tracking the target through an infrared optical device. The gunner must exert considerable motor control over the TOW launcher to maintain the sight at the target's center and make both gross and fine motor adjustments (Cartner, et. al, 1985).

These coordination and tracking skills needed by the TOW systems are not readily assessed by ASVAB. New Project A tests computerized psychomotor tasks and paper-and-pencil spatial tests have been developed to measure these kinds of abilities. This research was undertaken to see if an assignment procedure that includes measures of psychomotor and spatial ability would be likely to reduce training time and costs and/or result in more proficient TOW gunners.

METHOD

Sample

Subjects were 410 11H soldiers who participated in the Project A testing during the Longitudinal Validation of the Predictor Battery during FY 87. M-70 Qualification Scores were obtainable for 353 of these soldiers. A smaller subsample of 81 soldiers was used in some of the analyses. These were soldiers for whom additional performance data were available. These data included the number of attempts required to pass each of eight hands-on measures of critical tasks tested during training in addition to tracking.

Description of the Instruments

The instruments used in this research were developed to measure cognitive, psychomotor, perceptual and spatial abilities. A summary description is provided in Table 1. Additional information is provided below and in Peterson (1987).

(INSERT TABLE 1 ABOUT HERE)

Paper-and-Pencil Tests

Reasoning. This test measures the ability to identify a logical rule or pattern governing a series of figures and select the next figure in the series from five alternatives.

Object Rotation. This test measures the ability mentally to rotate two-dimensional figures or to manipulate their component parts. The examinee must indicate whether an item figure is the "same" (rotated) or "not same" (flipped) as the stimulus.

Assembling Objects. This test measures ability to visualize how component parts will fit together to form a completed object. Each test item presents a picture of several components. The examinee must select the one response out of four that shows the parts correctly put together.

Orientation. This test measures the ability to maintain one's perspective or bearing with respect to some object when it and its component parts have been rotated. Each item presents a non-upright picture within a frame containing a circle with a dot at the bottom. The

examinee mentally must rotate the frame to align the circle with the bottom of the picture. Then he must determine the position of the dot.

Maze. This test measures ability to scan a complex field visually and identify a particular pattern, configuration, or pathway within it. For each item, the examinee must decide which of four labeled (A to D) entrances to a maze leads to an exit.

Map. Subjects are presented with maps with landmarks (e.g., mess hall, campsite, lake) but no compass directions. Test items provide information about direction (e.g., "The forest is north of the campsite") which examinees use to determine direction to travel to another landmark.

Computer Tests

Overall. The computers used a specially designed response pedestal pictured in Figure 1. Two joysticks allow subjects to choose right or left hand. There are also horizontal and vertical slide controls, three response buttons (blue, white, and yellow), and two red response buttons. Four green "home" buttons ensure controlling the location of subjects' hands during stimulus presentation, controlling the onset of test items, and enabling assessment of both decision time and movement time components of reaction time.

(Insert Figure 1 about here)

Simple Reaction Time. The first five items of this test are practice items to enable the subject to get accustomed to the equipment and are not scored. Subjects start with hands on the HOME buttons (ready position). For each item, the subject must respond as quickly as possible when the word "YELLOW" appears by pressing the yellow response button. To eliminate extreme scores caused by extraneous factors, the fastest and slowest responses are deleted.

Choice Reaction Time. This test requires a choice response. Instead of the stimulus "YELLOW", the stimulus is either "BLUE" or "WHITE". When it appears, the subject must release the HOME button and press the correct color button.

Memory. This tests measures how quickly individuals can use information from short-term memory. A set of up to five letters briefly appears on the screen followed by presentation of a single probe letter. The task is to decide if the probe letter was contained in the original display and press the WHITE (yes) or BLUE (no) button.

One-Hand Tracking. This pursuit tracking task using a joystick requires the subject to keep crosshairs centered on a target box as it travels at a constant speed along a path with both horizontal and vertical segments. Speed of crosshairs and of targets, length of the path, and number of segments making up the path vary across trials.

Perceptual Speed & Accuracy. This tests requires examinees to compare rapidly two sets of up to 9 characters (letters, numbers, symbols, or real words) that are presented simultaneously. Subjects indicate pairs are "same" (WHITE button) or "different" (BLUE button).

Two-hand Tracking. This is another pursuit tracking test, but it is designed to measure multilimb (two-handed) coordination. The subject controls movement of the crosshairs using the vertical and horizontal slides instead of a joystick.

Number Memory. This tests the ability to do simple arithmetic operations quickly and accurately. Items consist of sequential presentation of numerical operations until an answer is presented. Subject presses a button to indicate if the solution is correct (WHITE) or not (BLUE).

Cannon Shoot. This test measures ability to hit a target moving at constant speed and direction by firing from a stationary position. The location of the cannon, speed and direction of the target, angle of target to cannon, and distance from cannon to impact point and from fire point to impact point vary across trials.

Target Identification Test. This test requires subjects to match target figures (simple renditions of military vehicles and aircraft) to one of three stimulus figures labelled BLUE, YELLOW, and WHITE, using the appropriate buttons

Target Shoot Test. At the start of a trial, the target box and crosshairs appear at different points on the screen. While the target moves about unpredictably, changing both speed and direction, the subject uses the joystick to center the crosshairs on the target and presses a RED button to "fire".

Criterion Variables

As is often the case in validation research in an applied or operational setting, the performance criterion was of primary concern here. An adequate criterion must be reliable, valid, and contain sufficient variance to differentiate between good and bad performance. The M-70 simulator is used for both training and evaluation during TOW gunner AIT, and the scores obtained from this simulator appear to meet these requirements (Hughes Aerospace Group, 1967). All TOW gunners must achieve a passing score on the M-70 simulator prior to graduation. For the subsample of cases for whom full training data was available, hands-on scores for eight critical tasks were provided as well.

The M-70 simulator requires tracking of a real target through an infrared optical device. The simulator provides noise effects comparable to a "live fire" situation. One drawback to the M-70 training exercise is that it does not require a soldier to track along both vertical and horizontal axes, i.e, on ground that is not level; instead, the training exercise requires a soldier to track a target that is

moving at a constant speed, in a horizontal trajectory, along a paved, level road. These particular M-70 data, therefore, may have limited generalizability to other more demanding tracking situations.

<u>M-70 Qualification Score</u>. The M-70 score is a measure of tracking behavior over a series of simulated TOW firings. The score for each firing reflects the probability of a hit. According to the Hughes Aerospace Group (1967), the score used to obtain the probability estimate is analogous to the standard deviation of the deviations of the TOW off of its ideal track during its trajectory. The scores on each trial are summed over all ten trials of the TOW exercise. The minimum score required to qualify as a TOW gunner is 550 out of a possible 1000.

<u>M-70 Rating</u>. These ratings are based on the qualification score. The four ratings have been assigned the values 0-3 for this research and are as follows: 0=Unqualified (score < 550), 1="2nd Class" (550-649), 2="1st Class" (650-749), 3="Expert" (750-1000). Using this criterion allowed the inclusion of an additional 17 soldiers who did not qualify and therefore did not have M-70 Qualification Scores.

<u>Hands-on Scores</u>. The TOW2 training program includes instruction in assembling, maintaining, and firing the TOW and in deployment and safety as well as tracking. Soldier trainees must demonstrate the ability to perform eight critical tasks vital to their job before being tested on the M-70 simulator. To demonstrate proficiency in these tasks, soldiers must receive a "Go" upon actual hands-on performance of the tasks. These tasks, as listed in the student progress chart, are:

1. Assemble the TOW2 launcher
2. Conduct system checkout procedures and preoperational inspection of the TOW2 launcher and encased missile.
3. Perform operator's maintenance on the TOW2 launcher.
4. Load, aim, and unload an encased missile.
5. Perform immediate action procedures for a TOW2 misfire.
6. Determine if a target is engageable.
7. Prepare an anti-armor range card (TOW).
8. Explain TOW firing limitations and characteristics.

The relationship between these important aspects of TOW gunnery performance, in addition to tracking, and Project A predictors was investigated as well. For this research, each task was scored as the number of attempts required to achieve a "Go" and an overall sum was computed. Four attempts were allowed on each task; failure to achieve a "Go" was scored as a "5" on tasks 1-7. All reported scores on task 8 were "1"; the few cases where task 8 was unscored were also considered to have scored a "1". Hands-on scores were available for the subsample only.

<u>Procedures</u>

<u>Predictor testing</u>. Soldiers were tested on the Project A measures in the Reception Station. Approximately half the subjects took the computer

tests first while others took the pencil-and-paper tests first. All subjects received the timed pencil-and-paper tests in the order indicated in Table 1. Subjects were able to ask questions after the instructions were read to them. The computerized tests were self-administered on Seequa Chameleon computers. All instructions were provided on the computer screen. A test administrator was present during testing to handle questions and/or equipment problems.

Criterion Testing. Scores were obtained from records and not subject to researcher control. Normal procedure for training in firing and tracking on the M-70 includes four days of training under different conditions (e.g., ranges, target speeds,night tracking, rapid fire). Qualification firing on the fifth day involves firing in the M-70 fitted with the M-80 blast simulator (Cartner, et. al., 1985).

Analyses

General descriptive statistics (means and standard deviations) were calculated for predictor and M-70 scores. Correlations between the predictor tests and criterion scores for the total sample and for the subsample were calculated as well.

RESULTS

Means and standard deviations for each of the tests used from the Project A Battery, the M-70 Qualification Score, and the ASVAB Aptitude Area CO (Combat) score for the total sample and for the subsample are provided in Tables 2 and 3 respectively. A comparison of the means in Tables 2 and 3 indicates that only Perceptual Speed and Accuracy Decision Time and Target Identification Decision Time means were significantly different between the two groups at the .05 level. Because two of the 33 comparisons would be expected to be significant by chance alone, no further effort was expended in group comparisons.

(Insert Tables 2 and 3 about here)

Table 4 contains the correlations between the Project A predictors and the criteria for both the sample and the subsample. The two criteria used for the sample were the M-70 Qualification score and the M-70 rating. The subsample also has correlations reported for the total Hands-on score. The Project A tests that significantly predict M-70 Qualification scores and ratings in both samples are One-Hand Tracking, Two-Hand Tracking, Reasoning and Orientation. In the total sample, the additional significant predictors of both criteria were Object Rotation and Map. In the subsample, the Maze test predicted all three criteria significantly. In general, the correlation coefficients are in the expected direction. Longer response times and greater tracking error distances are usually indicative of poorer performance.

(Insert Table 4 about here)

Another way to evaluate the potential impact of using Project A measures to choose soldiers to enter gunner training is to compare the performance of soldiers scoring high on the Project A measures with the total group. In this analysis, the total sample was divided into thirds based on the Total Tracking Score, the sum of standardized scores for One-hand and Two-hand tracking. The distribution of M-70 ratings was then computed for those scoring in the top third of the Total Tracking Score as well as for the entire sample. Figure 2 summarizes the comparison made between the two groups. In the total sample, 8.9 percent of the soldiers scored as Experts on their final M-70 Qualification Score; in the top third subsample, the percentage of Experts increased to 12.9 percent. Perhaps a more important difference is seen for those who qualified as 2nd Class. In the total sample, 61.0 percent scored as 2nd Class while in the group having "high" Project A Total Tracking scores only 48.3 percent fell in this lowest qualified category.

(Insert Figure 2 about here)

DISCUSSION

As in the case of the UCOFT research, this preliminary research into the effectiveness of Project A psychomotor and spatial tests in predicting success in TOW gunner training yields promising results. Using one or more of the Project A tests to choose Infantry basic trainees to be assigned to further training as TOW gunners should result in more proficient gunners at a relatively small added cost for testing.

Additional analyses need to be performed on the subsample where additional data is available on performance during training. These additional data include scores on individual Hands-On tasks as well as an initial M-70 score that is not as confounded with training and practice effects as is the M-70 Qualification Score. If additional training data could be collected, the effectiveness of the Project A tests in predicting failures could also be evaluated.

The next step in this research will be to determine which of the Project A tests could add both independent and significant validity to the assignment process for TOW gunners. Although the current computer and spatial test battery requires approximately two and one-half to three hours to administer, a reduced battery of two to four tests found to be sufficient for TOW gunners could be administered in less than one hour.

REFERENCES

Cartner, J. A., Strasel, H. C., Evans, K. L., Heller, F. H., & Tierney, T. (1985). TOW gunner selection. (ARI Research Note 85-72). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N. G. (Ed.) (1987) Development and field test of the Trial Battery for Project A. (ARI Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hughes Aerospace Group. Aeronautical Systems Division (1967). TOW scoring unit evaluation study. (Report No. TOW-T36). Culver City, CA: J. A. Scanlon.

Smith, E. P. & Graham, S. E. (1987) Validation of psychomotor and perceptual predictors of Armor Officer M-1 gunnery performance. (ARI Technical Report No. ___). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Table 1

Summary of Project A Predictor Measures

Test Name	Construct	Number of Items	Time Limit	Score	Reliability
Paper and Pencil Tests					
Reasoning Test	Induction	30	12 min.	No. correct	.87[a]
Object Rotation Test	Spatial Visualization: Rotation	90	7.5 min.	No. correct	.99
Orientation Test	Spatial Orientation	24	10 min.	No. correct	.89
Maze Test	Spatial Visualization: Scanning	24	5.5 min.	No. correct	.96
Map Test	Spatial Orientation	20	12 min.	No. correct	.90
Assembling Objects	Spatial Visualization: Rotation	32	26 min.	No. correct	.91
Computerized Tests					
Simple Reaction Time	Processing Efficiency	15	---	Decision time mean	.88
				Movement time mean	--
				% correct	.46
Choice Reaction Time	Processing Efficiency	30	---	Decision time mean	.97
				Movement time mean	--
				% correct	.57
Memory Test	Short Term Memory	36	---	Decision time mean	.96
				Movement time mean	--
				% correct	.60
Target Tracking 1	Psychomotor Precision	18	---	Mean log (distance + 1)	.98
Perceptual Speed and Accuracy	Perceptual speed and accuracy	36	---	Decision time mean	.94
				Movement time mean	--
				% correct	.65
Target Tracking 2	Two-hand coordination	18	---	Mean log (distance + 1)	.98
Number Memory Test	Number Operations	28	---	Final response time mean	.88
				Input response time mean	.95
				Pooled operation mean	.93[b]
				% correct	.59
Cannon Shoot	Movement judgment	36	---	Absolute time discrepancy Mean	.65
				Mean log (distance + 1) error	--
				% hits	--
Target Identification	Perceptual speed and accuracy	36	---	Decision time mean	.97
				Movement time mean	--
				% correct	.62
Target Shoot Test	Psychomotor precision	30	---	Mean log (distance + 1) error	.74
				Mean time to fire	.85
				% hits	--

[a] Split-half (odd-even) reliability with Spearman-Brown Correction for length based on Project A Concurrent Validation (CV) data (n = 9332 to 9345).
[b] Coefficient alpha reliability based on CV data.

Table 2

Total Sample Means and Standard Deviations for ASVAB CO Aptitude Area Composite, the Project A Tests and M70 Qualification Score

	N[a]	Mean[b]	S.D.
ASVAB			
CO Composite	406	113.5	12.4
Simple Reaction Time			
Percent Correct	409	98	06
Decision Time Mean	409	30.11	11.36
Movement Time Mean	409	25.64	16.61
Choice Reaction Time			
Percent Correct	410	98	04
Decision Time Mean	410	38.47	8.03
Movement Time Mean	410	24.29	6.90
Short Term Memory			
Percent Correct	406	89	07
Decision Time Mean	406	77.33	21.00
Movement Time Mean	406	34.08	11.66
One-hand Tracking			
Mean Log (Dist + 1) Error	408	2.66	.38
Perceptual Speed and Accuracy			
Percent Correct	407	86	09
Decision Time Mean	407	210.85	60.87
Movement Time Mean	408	30.81	10.63
Two-hand Tracking			
Mean Log (Dist +1) Error	408	3.30	.47
Number Memory			
Percent Correct	404	88	10
Initial Response Time Mean	404	13.20	42.2
Pooled Operations Time Mean	404	20.18	7.63
Final Response Time Mean	404	137.52	37.00
Cannon Shoot			
Absolute Time Descrepancy Mean	403	41.04	7.45
Target Identification			
Percent Correct	398	90	07
Decision Time Mean	398	154.74	51.81
Movement Time Mean	398	33.78	9.20
Target Shoot			
Mean Time To Fire	393	208.27	.22
Mean Log (Dist + Error)	393	2.08	45.36
Reasoning	390	20.7	5.1
Object Rotation	390	63.1	19.4
Orientation	390	13.0	6.5
Maze	390	18.4	4.3
Map	390	9.1	5.6
Assembling Objects	390	25.2	6.6
M70 Qualification Score	353	651.8	67.6

[a]Ns differ due to missing data

[b]Times are in hundredths of seconds

Table 3

Subsample Means and Standard Deviations for ASVAB CO Aptitude Area Composites, Project A Tests and M70 Qualification Score

	N[a]	Mean[b]	S.D.
ASVAB	79	111.4	11.4
CO Composite			
Simple Reaction Time			
Percent Correct	81	98	05
Decision Time Mean	81	28.79	6.31
Movement Time Mean	81	23.84	11.00
Choice Reaction Time			
Percent Correct	81	98	03
Decision Time Mean	81	37.29	6.28
Movement Time Mean	81	24.07	7.92
Short Term Memory			
Percent Correct	80	90	06
Decision Time Mean	80	74.00	16.54
Movement Time Mean	80	31.66	9.65
One-Hand Tracking			
Mean Log (Dist + 1) Error	80	2.69	.42
Perceptual Speed and Accuracy			
Percent Correct	80	85	10
Decision Time Mean	80	195.77	58.68
Movement Time Mean	80	28.80	8.55
Two-Hand Tracking			
Mean Log (Dist +1) Error	80	3.29	.48
Number Memory			
Percent Correct	79	88	10
Initial Response Time Mean	79	13.09	4.64
Pooled Operations Time Mean	79	18.74	7.07
Final Response Time Mean	79	132.50	32.48
Cannon Shoot			
Absolute Time Descrepancy Mean	77	40.84	6.90
Target Identification			
Percent Correct	76	89	08
Decision Time Mean	76	143.51	41.71
Movement Time Mean	76	32.91	8.64
Target Shoot			
Mean Time To Fire	72	209.30	.17
Mean Log (Dist + Error)	72	2.06	50.02
Reasoning	80	20.6	5.3
Object Rotation	80	63.0	19.2
Orientation	80	13.7	5.9
Maze	80	18.6	4.2
Map	80	9.1	5.3
Assembling Objects	80	26.1	5.4
M70 Qualification Score	81	640.6	70.7

[a]Ns differ due to missing data

[b]Times are in hundredths of seconds

Table 4

Zero-Order Correlations Between Predictor Variable and M-70 Criterion Scores

Predictors	Criteria				
	Total Sample		Subsample[c]		
	M70 Qual.[a] Scores	M70[b] Rating	M70 Qual. Scores	M70 Rating	Hands on Total
Simple Reaction Time					
Percent Correct	-.01	-.01	.04	.17	.07
Decision Time Mean	-.02	-.02	.10	.07	.13
Movement Time Mean	00	00	-.14	-.18	-.05
Choice Reaction Time					
Percent Correct	00	-.03	-.05	-.05	-.08
Decision Time Mean	.05	-.01	.19	.15	.01
Movement Time Mean	-.07	-.10*	-.12	-.19	-.03
Short Term Memory					
Percent Correct	-.07	.03	.06	.00	-.19
Decision Time Mean	.05	.00	.16	.20	.09
Movement Time Mean	-.02	-.08	-.24*	-.20	.03
Tracking 1					
Mean Log (Dist + 1) Error	-.18***	-.18***	-.41***	-.34**	.31**
Perceptual Speed and Accuracy					
Percent Correct	.01	-.02	.02	.01	-.18
Decision Time Mean	-.07	-.06	-.09	-.16	-.09
Movement Time Mean	.01	-.04	-.16	-.17	-.02
Tracking 2					
Mean Log (Dist +1) Error	-.28***	-.27***	-.40***	-.35**	.29**
Number Memory					
Percent Correct	.05	.05	.19	.17	-.21
Initial Response Time Mean	-.06	-.08	-.11	-.11	.05
Pooled Operations Time Mean	-.02	-.03	.01	-.01	.08
Final Response Time Mean	-.01	-.02	-.02	.00	.07
Cannon Shoot					
Absolute Time Descrepancy Mean	-.15**	-.13*	-.22	-.24*	.12
Target Identification					
Percent Correct	-.02	-.04	-.10	-.09	.05
Decision Time Mean	-.18***	-.16**	-.18	-.19	.04
Movement Time Mean	.02	-.02	-.04	-.01	.04
Target Shoot					
Mean Time To Fire	-.11	-.16**	-.40***	-.37**	.20
Mean Log (Dist + Error)	-.10	-.12*	-.22	-.16	.27*
Reasoning No. Correct	.14*	.16**	.28*	.29**	-.24*
Object Rotation No. Correct	.17**	.20***	.12	.15	-.20
Orientation No. Correct	.23***	.28***	.33**	.37***	-.34***
Maze No. Correct	.08	.14**	.32**	.29**	-.38**
Map No. Correct	.16**	.21***	.20	.15	-.14
Assembling Objects	.06	.11*	.17	.11	-.31**

[a]ns range from 338-353
[b]ns range from 355-370
[c]ns range from 72 - 81
*p < .05; **p < .01; ***p < .001

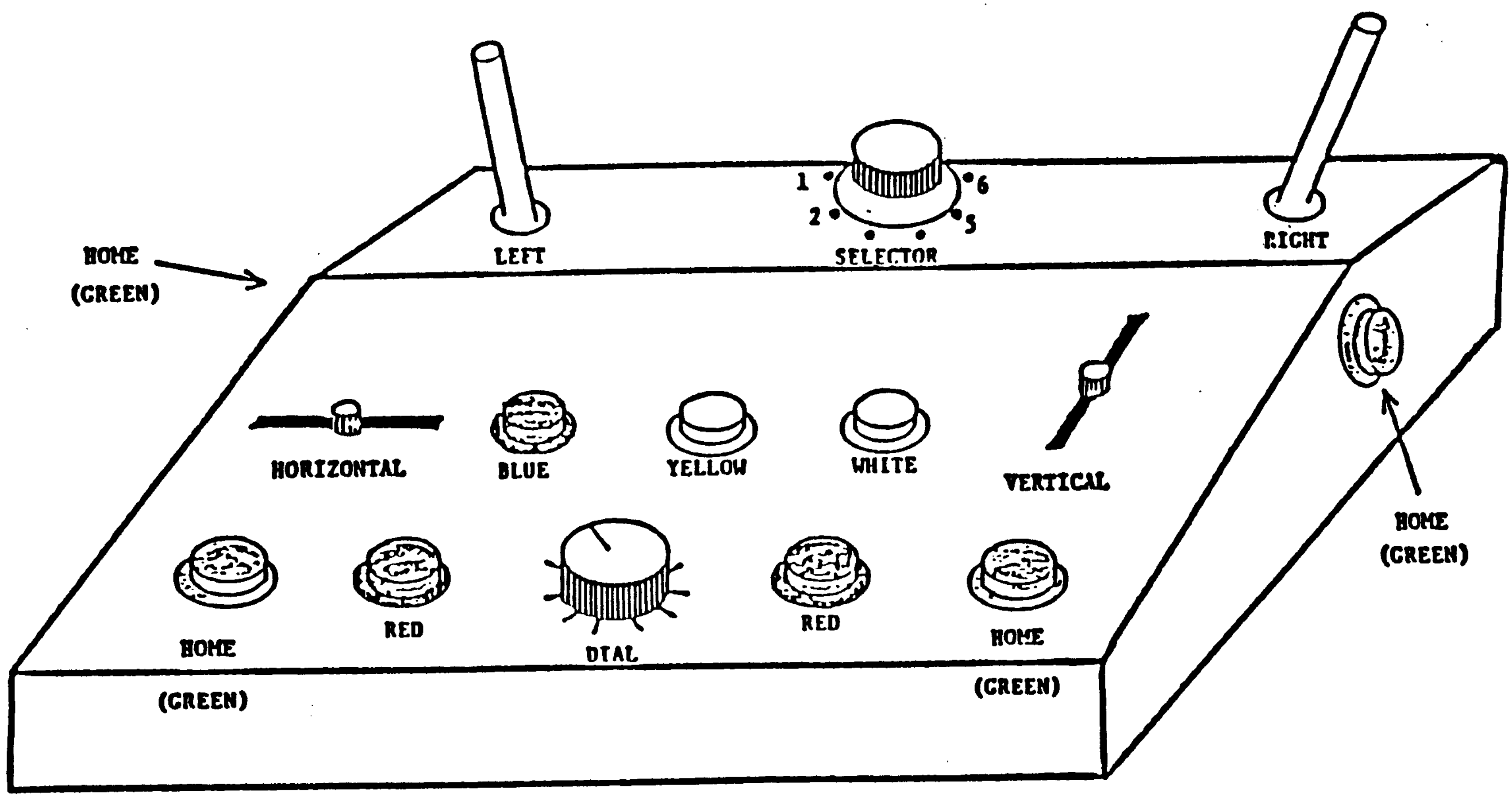


Figure 1. Response pedestal for computerized tests.

Improving TOW Gunnery
Screening for Psychomotor Skills

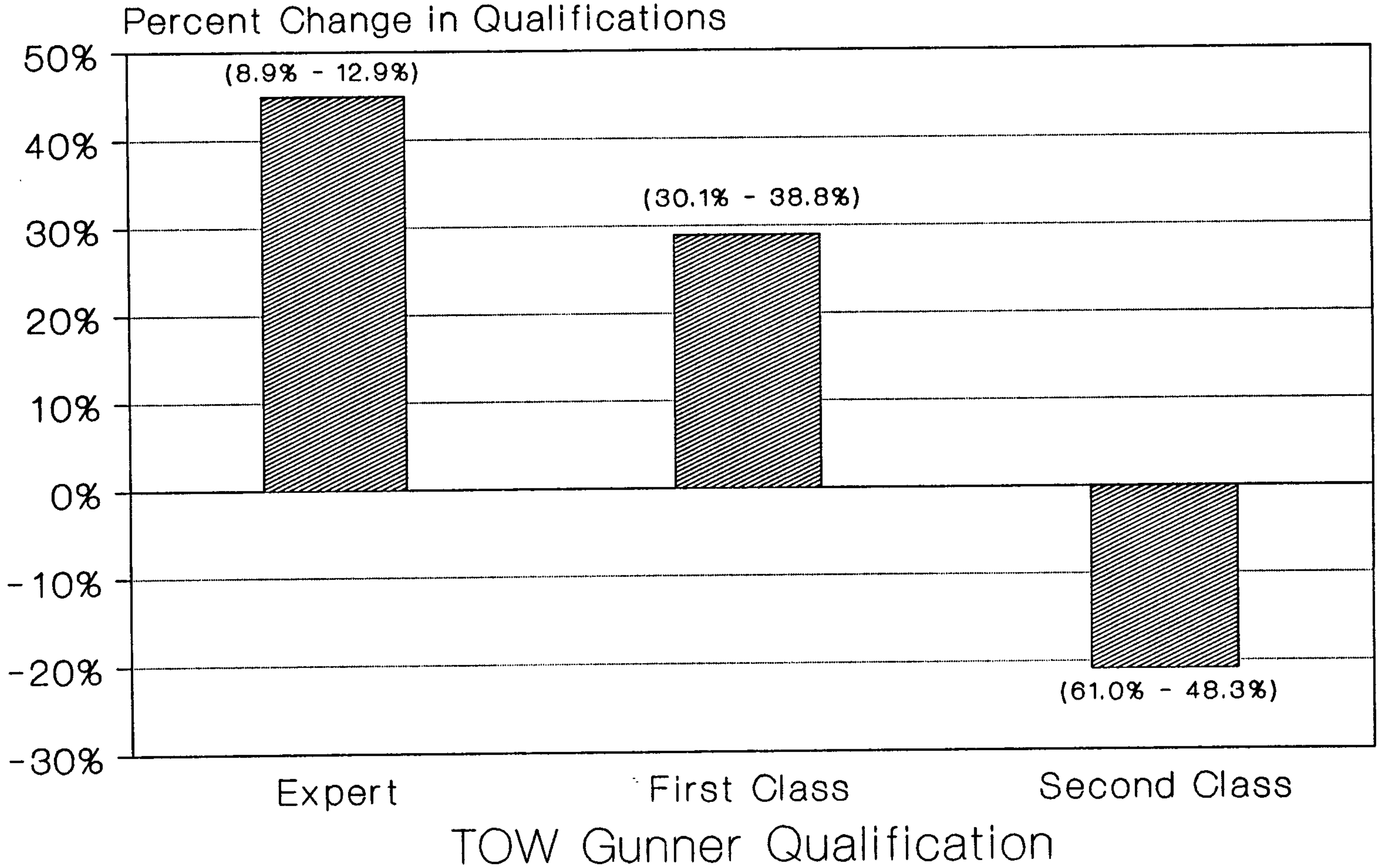


Figure 2

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-1

Relationship Between Project A Psychomotor and Spatial Tests and TOW2 Gunnery Performance: A Preliminary Investigation

Frances Grafton, Mark Y. Czarnolewski and Elizabeth P. Smith

January 1989

REVIEWED BY:

Clinton B. Walker
Team Chief, Predictor Development, SCTA

APPROVED BY:

David W. Witter
Assistant Director, Manpower and Personnel Research Laboratory

CLEARED BY:

for Newell K. Eaton, Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Acknowledgments

I would like to express my appreciation to the following people for providing assistance with this project:

Army Research Institute -- Alexandria

Mr. John Miles - Systems Research Laboratory
Major Alan Brauer - Systems Research Laboratory
Ms. Frances Grafton - Manpower and Personnel Research Laboratory
Dr. Michael Rumsey - Manpower and Personnel Research Laboratory

Army Research Institute -- Fort Lewis Field Unit

Dr. John Whittenburg, ARI Liaison ADEA

Fort Lewis

ADEA:

Major Jesse Johnson - Force Development Division, Training Branch
Major Jim Carter - Chief, TEO
Mr. Robert Gorak - Assistant Director, TEO
Mr. Robert Fuchs - Scientific Advisor

Anti-Armor Theater:

NCOIC, SFC: Joseph Guillen
SP4: Warren Nicholson

Allen Corporation of America

Mr. Randy Jones
Dr. Timothy O'Keefe
Dr. Jose Guerrier

Human Resources Research Organization (HumRRO)

Mr. Jim Harris
Mr. Michael Dunn
Ms. Mary Duffey
Ms. Jody Toquam
Mr. Joe Hauth

American Institutes for Research (AIR)

Ms. Winnie Young

Special thanks to Dr. Lawrence Hanser who served as Technical Area Chief, to Frances Grafton, Sidney Sachs and Richard Maisano for their constructive editorial comments, and to Betty Shelly whose secretarial skills were only matched by her aplomb and professionalism. Also, thanks to Dr. David Promisel, who suggested the possibility of doing this project.

Relationship Between Project A Psychomotor and Spatial Tests and TOW2 Gunnery Performance: A Preliminary Investigation

INTRODUCTION

Within the Army, matching the right person to the right job continues to be an important issue. Ongoing research efforts attempt to optimize person-to-job matches. One such major effort is the Army's Project A: Improving the Selection, Classification and Utilization of Army Enlisted Personnel. One of Project A's principal goals is to improve traditional methods and add new ones to enable measurement of a greater range of abilities required for successful soldier performance (Eaton, Goer, Harris, & Zook, 1984). This work has resulted in development of a computerized and pencil-and-paper test battery of temperament, cognitive, psychomotor, perceptual and spatial variables which are expected to predict performance in a diversity of Military Occupational Specialties (MOS). Early results indicate that the Project A battery has incremental validity in predicting enlisted performance over that obtained with the Armed Services Vocational Aptitude Battery (ASVAB) alone, the single current selection/classification tool.

As work on Project A has progressed, it has become apparent that there are special applications for which parts of the battery have considerable potential. Among the most apparent is the use of psychomotor and spatial tests for the selection of gunners. The tests have already been researched with M-1 tank gunners using the Unit Conduct of Fire Trainer (UCOFT) (Smith & Graham, 1987). The high correlation between the tests and UCOFT performance ($\underline{R}$ = .76) suggests that similar results with other gunnery specialties are possible. The present effort extends research on the validity of the Project A tests for assigning basic trainees in Infantry into Advanced Individual Training (AIT) for MOS 11H (TOW2 gunner). The work was undertaken at the request of the Commanding General of the Training and Doctrine Command (TRADOC).

Background

The advent of anti-armor systems that fire wire-guided missiles for Infantry use brought new considerations to the issues of personnel selection and training (Cartner, Strasel, Evans, Heller, & Tierney, 1985). Both the heavy anti-tank TOW2 (Tube-launched, Optically-tracked, Wire-guided) and weapons of the "fire and forget" category such as the bazooka and M-1 tank require abilities of target acquisition and identification, prediction of movement direction, and judgment of when to fire. The TOW2, however, also requires that the gunner maintain the missile on target until the missile hits, i.e., continuous sight alignment. The ground mount TOW2, the system of interest in this research, requires the gunner to manipulate knobs with both hands to move along a vertical axis while visually tracking the target through an infrared optical device. The TOW2 is swiveled on its ground mount to move the missile left or right. The gunner must exert considerable motor control over the TOW2 launcher to maintain the sight at the target's center and make both gross and fine motor adjustments (Cartner, et. al, 1985).

These coordination and tracking skills needed by the TOW2 system are not readily assessed by ASVAB. New Project A computerized psychomotor tasks and paper-and-pencil spatial tests have been developed to measure these kinds of abilities. This research was undertaken to see if an assignment procedure that includes measures of psychomotor and spatial ability would be likely to reduce training time and costs and/or result in more proficient TOW2 gunners.

METHOD

Sample

Subjects were 410 11H soldiers who participated in the Project A testing during the Longitudinal Validation of the Predictor Battery during FY 87. M-70 Qualification Scores were obtained for 353 of these soldiers. A smaller subsample of 81 soldiers was used in some of the analyses. These were soldiers for whom additional performance data were available. These additional data included the number of attempts required to pass each of eight hands-on measures of critical tasks tested during training and final M-70 scores for soldiers who did not graduate form AIT.

Description of the Instruments

The instruments used in this research were developed to measure cognitive, psychomotor, perceptual and spatial abilities. A summary description is provided in Table 1. Additional information is provided below and in Peterson (1987).

(Insert Table 1 about here)

Paper-and-Pencil Tests

Spatial Reasoning. This test measures the ability to identify a logical rule or pattern governing a series of figures and select the next figure in the series from five alternatives.

Object Rotation. This test measures the ability mentally to rotate two-dimensional figures or to manipulate their component parts. The examinee must indicate whether an item figure is the "same" (rotated) or "not same" (flipped) as the stimulus.

Assembling Objects. This test measures the ability to visualize how component parts will fit together to form a completed object. Each test item presents a picture of several components. The examinee must select the one response out of four that shows the parts correctly put together.

Orientation. This test measures the ability to maintain one's perspective or bearing with respect to some object when it and its component parts have been rotated. Each item presents a non-upright picture within a frame containing a circle with a dot at the bottom. The examinee mentally must rotate the frame to align the circle with the bottom of the picture. Then he must determine the position of the dot.

Maze. This test measures the ability to scan a complex field visually and identify a particular pattern, configuration, or pathway within it. For each item, the examinee must decide which of four labeled (A to D) entrances to a maze leads to an exit.

Map. Subjects are presented with maps containing landmarks (e.g., mess hall, campsite, lake) but no legend to identify compass directions. Test items, however, provide information about direction (e.g., "The forest is north of the campsite") which examinees use to determine direction to travel of another landmark.

Computer Tests

Overall. The computers used a specially designed response pedestal pictured in Figure 1. Two joysticks allow subjects to choose right or left hand. There are also horizontal and vertical slide controls, three response buttons (blue, white, and yellow), and two red response buttons. Four green "home" buttons ensure controlling the location of subjects' hands during stimulus presentation, controlling the onset of test items, and enabling assessment of both decision time and movement time components of reaction time.

(Insert Figure 1 about here)

Simple Reaction Time. The first five items of this test are practice items to enable the subject to get accustomed to the equipment and are not scored. Subjects start with hands on the HOME buttons (ready position). For each item, the subject must respond as quickly as possible when the word "YELLOW" appears by pressing the yellow response button. To eliminate extreme scores caused by extraneous factors, the fastest and slowest responses are deleted.

Choice Reaction Time. This test requires a choice response. Instead of the stimulus "YELLOW", the stimulus is either "BLUE" or "WHITE". When it appears, the subject must release the HOME button and press the correct color button.

Memory. This test measures how quickly individuals can use information from short-term memory. A set of up to five letters briefly appears on the screen followed by presentation of a single probe letter. The task is to decide if the probe letter was contained in the original display and press the WHITE (yes) or BLUE (no) button.

One-Hand Tracking. This pursuit tracking task, using a joystick, requires the subject to keep crosshairs centered on a target box as it travels at a constant speed along a path with both horizontal and vertical segments. Speed of crosshairs and of targets, length of the path, and number of segments making up the path vary across trials.

Perceptual Speed & Accuracy. This test requires examinees to compare two simultaneously-presented sets of stimulus arrays with up to 9 characters each (letters, numbers, symbols, or real words). Subjects indicate whether the two sets are "same" (WHITE button) or "different" (BLUE button).

Two-hand Tracking. This is another pursuit tracking test, but it is designed to measure multilimb (two-handed) coordination. The subject controls movement of the crosshairs using the vertical and horizontal slides instead of a joystick.

Number Memory. This test measures the ability to do simple arithmetic operations quickly and accurately. Items consist of sequential presentation of numerical operations until an answer is presented. Subject presses a button to indicate if the answer is correct (WHITE) or not (BLUE).

Cannon Shoot. This test measures the ability to hit a target moving at constant speed and direction by firing from a stationary position. The location of the cannon, speed and direction of the target, angle of target to cannon, and distance from cannon to impact point and from fire point, to impact point vary across trials.

Target Identification Test. This test requires subjects to match target figures (simple renditions of military vehicles and aircraft) to one of three stimulus figures labelled BLUE, YELLOW, and WHITE, using the appropriate buttons.

Target Shoot Test. At the start of a trial, the target box and crosshairs appear at different points on the screen. While the target moves about unpredictably, changing both speed and direction, the subject uses the joystick to center the crosshairs on the target and presses a RED button to "fire" when the crosshairs rendezvous with the target box.

Criterion Variables

As is often the case in validation research in an applied or operational setting, the performance criterion was of primary concern here. An adequate criterion must be reliable, valid, and contain sufficient variance to differentiate between good and bad performance. The M-70 simulator is used for both training and evaluation during TOW2 gunner AIT, and the scores obtained from this simulator appear to meet these requirements (Hughes Aerospace Group, 1967). All TOW2 gunners must achieve a passing score on the M-70 simulator prior to graduation. Again, for the subsample of cases for whom full training data was available, hands-on scores for eight critical tasks were provided as well.

The M-70 simulator requires tracking of a real target through an infrared optical device. The simulator provides noise effects comparable to a "live fire" situation. One drawback to the M-70 training exercise is that it does not require a soldier to track along both vertical and horizontal axes, i.e, on ground that is not level; instead, the training exercise requires a soldier to track a target that is moving at a constant speed, in a horizontal trajectory, along a paved, level road. These particular M-70 data, therefore, may have limited generalizability to other more demanding tracking situations.

M-70 Qualification Score. The M-70 score is a measure of tracking behavior over a series of simulated TOW2 firings. The score for each firing reflects

the probability of a hit. According to the Hughes Aerospace Group (1967), the score used to obtain the probability estimate is analogous to the standard deviation of the deviations of the TOW2 off of its ideal track during its trajectory. The scores on each trial are summed over all ten trials of the TOW2 exercise. The minimum M-70 score required to qualify as a TOW2 gunner is 550 out of a possible 1000.

M-70 Rating. These ratings are based on the qualification score. The four ratings have been assigned the values 0-3 for this research and are as follows: 0=Unqualified, 1="2nd Class" (550-649), 2="1st Class" (650-749), 3="Expert" (750-1000).

Hands-on Scores. The TOW2 training program includes instruction in assembling, maintaining, and firing the TOW2 and in deployment and safety as well as tracking. Soldier trainees must demonstrate the ability to perform eight critical tasks vital to their job as well as achieve a passing score on the M-70 simulator before they can graduate. These tasks, as listed in the student progress chart, are:

1. Assemble the TOW2 launcher
2. Conduct system checkout procedures and preoperational inspection of the TOW2 launcher and encased missile.
3. Perform operator's maintenance on the TOW2 launcher.
4. Load, aim, and unload an encased missile.
5. Perform immediate action procedures for a TOW2 misfire.
6. Determine if a target is engageable.
7. Prepare an anti-armor range card (TOW2).
8. Explain TOW2 firing limitations and characteristics.

The relationship between these important aspects of TOW2 gunnery performance, in addition to tracking, and Project A predictors was investigated as well. For this research, each task was scored as the number of attempts required to achieve a "Go" and an overall sum was computed. Four attempts were allowed on each task; failure to achieve a "Go" was scored as a "5" on tasks 1-7. All reported scores on task 8 were "1"; the few cases where task 8 was unscored were also considered to have scored a "1". Hands-on scores were available for the subsample only.

Procedures

Predictor testing. Soldiers were tested on the Project A measures in the Reception Station prior to being selected for training in MOS 11H. Approximately half the subjects took the computer tests first while others took the pencil-and-paper spatial tests first. All subjects received the timed pencil-and-paper tests in the order indicated in Table 1. Subjects were able to ask questions after the instructions were read to them. The computerized tests were self-administered on Seequa Chameleon computers. All instructions were provided on the computer screen. A test administrator was present during testing to handle questions and/or equipment problems.

Criterion Testing. Scores were obtained from records and not subject to researcher control. Normal procedure for training in firing and tracking on the M-70 includes four days of training under different conditions (e.g., ranges, target speeds, night tracking, rapid fire). Qualification firing on the fifth day involves firing in the M-70 fitted with the M-80 blast simulator (Cartner, et. al., 1985).

Analyses

General descriptive statistics (means and standard deviations) were calculated for predictor and M-70 scores. Correlations between the predictor tests and criterion scores for the total sample and for the subsample were calculated as well.

RESULTS

Means and standard deviations for each of the tests used from the Project A Battery, the M-70 Qualification Score, and the ASVAB Aptitude Area CO (Combat) score for the total sample and for the subsample are provided in Tables 2 and 3 respectively. A comparison of the means in Tables 2 and 3 indicates that only Perceptual Speed and Accuracy Decision Time and Target Identification Decision Time means were significantly different between the two groups at the .05 level. Because two of the 33 comparisons would be expected to be significant by chance alone, no further effort was expended in group comparisons.

(Insert Tables 2 and 3 about here)

Table 4 contains the correlations between the Project A predictors and the criteria for both the total sample and the subsample. The two criteria used for the sample were the M-70 Qualification score and the M-70 rating. The subsample also has correlations reported for the total Hands-on score. The Project A tests that significantly predict M-70 Qualification scores and ratings in both samples are One-Hand Tracking, Two-Hand Tracking, Spatial Reasoning and Orientation. In the total sample, the additional significant predictors of both criteria were Object Rotation and Map. In the subsample, the Maze test predicted all three criteria significantly. In general, the correlation coefficients are in the expected direction. Longer response times and greater tracking error distances are usually indicative of poorer performance.

(Insert Table 4 about here)

As can be seen in Table 4 there is an apparent difference in magnitude of the correlations between the two data sets, with the correlations in the smaller data set being larger than in the full data set. The differences between correlations in the two data sets were found to be significant ($p \leq .05$) for One-handed Tracking, Target Shoot, and the Maze tests (McNemar, 1969). Table 5 presents the correlations with M-70 Qualification Scores, Zs and the standard error of the difference between the Zs for these tests.

(Insert Table 5 about here)

The small data set differed slightly from the total data set. As mentioned earlier, M-70 scores were available for all soldiers in the small data set including the three soldiers who failed to qualify. For the rest of the total data set, however, no M-70 scores were received for soldiers who failed the course. There were a total of 19 failures (including the three soldiers from the subsample). All M-70 Qualification Scores for these soldiers were set to missing, and they were included only in the analysis of the M-70 rating scores. In analyes of the subsample, the reported M-70 scores were retained for the three course failures.

Among the three failures in the subset was a significant outlier whose final M-70 score was 312. The mean M-70 score for the three failures was 502 as compared to 646 for the other 78 soldiers in the subsample. In order to determine the possible effect the extremely low M-70 score might have had upon the correlations in the subsample the lowest M-70 score (312) was recoded to the next lowest score(550) and the correlations were recomputed. Table 6 compares the correlations computed using the original scores, the scores with the outlier recoded and the total sample M-70 Qualification scores. With the exception of the Object Rotation test all of the correlations decreased in the subsample with the outlier on the M-70 score recoded to the next higher score. The correlations shown in Table 5 as being significantly different from the total sample are no longer significantly different using the recoded score (see Table 6).

For the total sample, however, M-70 qualification scores were available only for those soldiers who qualified as TOW2 gunners. The M-70 Qualification score analyses were therefore "restricted" to graduates only since no scores were available for non-graduates. An additional analysis was performed on the total sample after assigning all non-graduates an M-70 score of 549 (the highest possible non-qualifying score.) The results of this analysis are presented in Table 7 for those predictors included in Table 6. As can be sseen in Table 7, the correlation for the total sample increase for all predictors when scores were assigned to the non-graduates.

The results of the two additional analyses show that the two sets of correlations are much more similar when adjustments are made for known discrepancies in the data sets (extreme outlier in the subsample and assigning failing M-70 scores to course failures in the total sample.) If additional data from student progress charts could be collected, the effect of outliers would be reduced and actual final M-70 scores would provide a more acceptable criterion.

Another way to evaluate the potential impact of using Project A measures to choose soldiers to enter gunner training is to compare the performance of soldiers scoring high on the Project A measures with the total group. In this analysis, the total sample was divided into thirds based on the Total Tracking Score, the sum of standardized scores for One-hand and Two-hand tracking. The distribution of M-70 ratings was then computed for those scoring in the top third of the Total Tracking Score as well as for the entire sample. Figure 2 summarizes the comparison made between the two groups. In the total sample,

8.9 percent of the soldiers scored as Experts on their final M-70 Qualification Score; in the top third subsample, the percentage of Experts increased to 12.9 percent. Perhaps a more important difference is seen for those who qualified as 2nd Class. In the total sample, 61.0 percent scored as 2nd Class while in the group having "high" Project A Total Tracking scores only 48.3 percent fell in this lowest qualified category.

(Insert Figure 2 about here)

DISCUSSION

As in the case of the UCOFT research, this preliminary research into the effectiveness of Project A psychomotor and spatial tests in predicting success in TOW2 gunner training yields promising results. Using one or more of the Project A tests to choose Infantry basic trainees to be assigned to further training as TOW2 gunners should result in more proficient gunners at a relatively small added cost for testing.

Additional analyses need to be performed on the subsample where additional data is available on performance during training. These additional data include scores on individual Hands-On tasks as well as an initial M-70 score that is not as confounded with training and practice effects as is the M-70 Qualification Score. If additional training data could be collected, the effectiveness of the Project A tests in predicting failures could also be evaluated.

The next step in this research will be to determine which of the Project A tests could add both independent and significant validity to the current assignment process for TOW2 gunners. Although the full computer and spatial test battery requires approximately two and one-half to three hours to administer, a reduced battery of two to four tests found to be sufficient for TOW2 gunners could be administered in less than one hour.

REFERENCES

Cartner, J. A., Strasel, H. C., Evans, K. L., Heller, F. H., & Tierney, T. (1985). TOW2 gunner selection. (ARI Research Note 85-72). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hughes Aerospace Group. Aeronautical Systems Division (1967). TOW2 scoring unit evaluation study. (Report No. TOW2-T36). Culver City, CA: J. A. Scanlon.

McNemar, Q. (1969). Psychological statistics (4th Edition). New York: John Wiley and Sons.

Peterson, N. G. (Ed.) (1987). Development and field test of the Trial Battery for Project A. (ARI Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith, E. P. & Graham, S. E. (1987) Validation of psychomotor and perceptual predictors of Armor Officer M-1 gunnery performance. (ARI Technical Report No. ___). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Table 1

Summary of Project A Predictor Measures

Test Name	Construct	Number of Items	Time Limit	Score	Reliability
Paper and Pencil Tests					
Reasoning Test	Induction	30	12 min.	No. correct	.87[a]
Object Rotation Test	Spatial Visualization: Rotation	90	7.5 min.	No. correct	.99
Orientation Test	Spatial Orientation	24	10 min.	No. correct	.89
Maze Test	Spatial Visualization: Scanning	24	5.5 min.	No. correct	.96
Map Test	Spatial Orientation	20	12 min.	No. correct	.90
Assembling Objects	Spatial Visualization: Rotation	32	26 min.	No. correct	.91
Computerized Tests					
Simple Reaction Time	Processing Efficiency	15	---	Decision time mean Movement time mean % correct	.88 -- .46
Choice Reaction Time	Processing Efficiency	30	---	Decision time mean Movement time mean % correct	.97 -- .57
Memory Test	Short Term Memory	36	---	Decision time mean Movement time mean % correct	.96 -- .60
Target Tracking 1	Psychomotor Precision	18	---	Mean log (distance + 1)	.98
Perceptual Speed and Accuracy	Perceptual Speed and Accuracy	36	---	Decision time mean Movement time mean % correct	.94 -- .65
Target Tracking 2	Two-hand coordination	18	---	Mean log (distance + 1)	.98
Number Memory Test	Number Operations	28	---	Final response time mean Input response time mean Pooled operation mean % correct	.88 .95 .93[b] .59
Cannon Shoot	Movement judgment	36	---	Absolute time discrepancy Mean Mean log (distance + 1) error % hits	.65 -- --
Target Identification	Perceptual speed and accuracy	36	---	Decision time mean Movement time mean % correct	.97 -- .62
Target Shoot Test	Psychomotor precision	30	---	Mean log (distance + 1) error Mean time to fire % hits	.74 .85 --

[a]Split-half (odd-even) reliability with Spearman-Brown Correction for length based on Project A Concurrent Validation (CV) data ($\underline{n}$ = 9332 to 9345).
[b]Coefficient alpha reliability based on CV data.

Table 2

Total Sample Means and Standard Deviations for ASVAB CO Aptitude Area Composite, the Project A Tests and M70 Qualification Score

	N[a]	Mean[b]	S.D.
ASVAB			
CO Composite	409	113.5	12.4
Simple Reaction Time			
Percent Correct	409	97.52	5.74
Decision Time Mean	409	30.11	11.36
Movement Time Mean	409	25.64	16.61
Choice Reaction Time			
Percent Correct	410	98.03	3.66
Decision Time Mean	410	38.47	8.03
Movement Time Mean	410	24.29	6.90
Short Term Memory			
Percent Correct	406	89.36	6.59
Decision Time Mean	406	77.33	21.00
Movement Time Mean	406	34.08	11.66
One-hand Tracking			
Mean Log (Dist + 1) Error	408	2.66	.38
Perceptual Speed and Accuracy			
Percent Correct	407	85.65	8.69
Decision Time Mean	407	210.85	60.87
Movement Time Mean	407	30.81	10.63
Two-hand Tracking			
Mean Log (Dist +1) Error	408	3.30	.47
Number Memory			
Percent Correct	404	99.12	9.57
Initial Response Time Mean	404	13.20	4.22
Pooled Operations Time Mean	404	20.18	7.63
Final Response Time Mean	404	137.52	37.00
Cannon Shoot			
Absolute Time Descrepancy Mean	403	41.04	7.45
Target Identification			
Percent Correct	398	90.06	7.20
Decision Time Mean	398	154.74	51.81
Movement Time Mean	398	33.78	9.20
Target Shoot			
Mean Time To Fire	393	208.27	45.36
Mean Log (Dist + Error)	393	2.08	.22
Reasoning	389	20.67	5.10
Object Rotation	393	63.06	19.43
Orientation	389	12.98	6.47
Maze	386	18.41	4.30
Map	386	9.08	5.66
Assembling Objects	386	25.18	6.61
M70 Qualification Score	353	651.81	67.56

[a]Ns differ due to missing data

[b]Times are in hundredths of seconds

Table 3

Subsample Means and Standard Deviations for ASVAB CO Aptitude Area Composites, Project A Tests and M70 Qualification Score

	N[a]	Mean[b]	S.D.
ASVAB			
CO Composite	79	111.40	11.44
Simple Reaction Time			
Percent Correct	81	97.92	5.30
Decision Time Mean	81	28.79	6.31
Movement Time Mean	81	23.84	11.00
Choice Reaction Time			
Percent Correct	81	98.22	2.72
Decision Time Mean	81	37.29	6.28
Movement Time Mean	81	24.07	7.92
Short Term Memory			
Percent Correct	80	89.72	6.17
Decision Time Mean	80	74.00	16.54
Movement Time Mean	80	31.66	9.65
One-Hand Tracking			
Mean Log (Dist + 1) Error	80	2.69	.42
Perceptual Speed and Accuracy			
Percent Correct	80	84.55	9.79
Decision Time Mean	80	195.77	58.68
Movement Time Mean	80	28.80	8.55
Two-Hand Tracking			
Mean Log (Dist +1) Error	80	3.28	.48
Number Memory			
Percent Correct	79	87.81	9.84
Initial Response Time Mean	79	13.09	4.64
Pooled Operations Time Mean	79	18.74	7.07
Final Response Time Mean	79	132.50	32.48
Cannon Shoot			
Absolute Time Descrepancy Mean	77	40.84	6.90
Target Identification			
Percent Correct	76	88.51	7.57
Decision Time Mean	76	143.51	41.71
Movement Time Mean	76	32.91	8.64
Target Shoot			
Mean Time To Fire	72	209.30	50.02
Mean Log (Dist + Error)	72	2.06	.17
Reasoning	80	20.66	5.33
Object Rotation	80	62.98	19.21
Orientation	80	13.68	5.93
Maze	80	18.60	4.19
Map	80	9.11	5.27
Assembling Objects	80	26.06	5.42
M70 Qualification Score	81	640.61	70.67

[a]Ns differ due to missing data

[b]Times are in hundredths of seconds

Table 4

Zero-Order Correlations Between Predictor Variable and M-70 Criterion Scores

Predictors	Criteria				
	Total Sample		Subsample[c]		
	M70 Qual.[a] Scores	M70[b] Rating	M70 Qual. Scores	M70 Rating	Hands on Total
Simple Reaction Time					
Percent Correct	-.01	-.01	.04	.17	.07
Decision Time Mean	-.02	-.02	.10	.07	.13
Movement Time Mean	.00	.00	-.14	-.18	-.05
Choice Reaction Time					
Percent Correct	-.00	-.03	-.05	-.05	-.08
Decision Time Mean	.05	-.01	.19	.15	.01
Movement Time Mean	-.07	-.10*	-.12	-.19	-.03
Short Term Memory					
Percent Correct	-.07	-.03	.06	.00	-.19
Decision Time Mean	.05	-.00	.16	.20	.09
Movement Time Mean	-.02	-.08	-.24*	-.20	.03
Tracking 1					
Mean Log (Dist + 1) Error	-.18***	-.18***	-.41***	-.34**	.31**
Perceptual Speed and Accuracy					
Percent Correct	.01	-.02	.02	.01	-.18
Decision Time Mean	-.07	-.06	-.09	-.16	-.09
Movement Time Mean	.01	-.04	-.16	-.17	-.02
Tracking 2					
Mean Log (Dist +1) Error	-.28***	-.27***	-.40***	-.35**	.29**
Number Memory					
Percent Correct	-.05	-.05	.19	.17	-.21
Initial Response Time Mean	-.06	-.08	-.11	-.11	.05
Pooled Operations Time Mean	-.02	-.02	.01	-.01	.08
Final Response Time Mean	-.01	-.03	-.02	-.00	.07
Cannon Shoot					
Absolute Time Descrepancy Mean	-.15**	-.13*	-.22	-.24*	.12
Target Identification					
Percent Correct	-.02	-.04	-.10	-.09	.05
Decision Time Mean	-.18***	-.16**	-.18	-.19	.04
Movement Time Mean	.02	-.02	-.04	-.01	.04
Target Shoot					
Mean Time To Fire	-.11	-.16**	-.40***	-.37**	.20
Mean Log (Dist + Error)	-.10	-.12*	-.22	-.16	.27*
Reasoning No. Correct	.14*	.16**	.28*	.29**	-.24*
Object Rotation No. Correct	.17**	.20***	.12	.15	-.20
Orientation No. Correct	.23***	.28***	.33**	.37***	-.34***
Maze No. Correct	.08	.14**	.32**	.29**	-.38**
Map No. Correct	.16**	.21***	.20	.15	-.14
Assembling Objects	.07	.11*	.17	.11	-.31**

[a]ns range from 338-353
[b]ns range from 355-370
[c]ns range from 72 - 81
*p < .05; **p < .01; ***p < .001

Table 5

Significant Differences in Correlations between Project A Tests and M-70 Qualification Scores for the Two Data Sets

Test	Total r	Subsample r	z	SE_{diff}
Tracking 1	.18	.44	2.21	.129
Target Shoot	.11	.40	2.69	.135
Maze	.08	.32	2.10	.130

Table 6

Comparison of Correlations Obtained Using Original and Recoded Data in the Subsample

Predictors	Total Sample M-70 Qual	Original Subsample	Recoded Subsample	z[a]
Tracking				
Tracking Composite	-.25***	-.43***	-.33***	
One-hand Tracking	-.18***	-.41***	-.30***	1.32
Two-hand Tracking	-.28***	-.40***	-.33***	
Target Shoot				
Time to Fire	-.11*	-.40***	-.30***	1.84
Distance	-.10	-.22	-.14	
Cannon Shoot				
Absolute Mean Disc. Time	-.15**	-.22	-.19	
Spatial Tests				
Spatial Reasoning	.14*	.28*	.25*	
Object Rotation	.17**	.12	.15	
Orientation	.23***	.33**	.31**	
Maze	.08	.32**	.26*	1.39
Map	.16**	.20	.19	
Assembling Objects	.06	.17	.11	

*p<_.05; **p<_.01; ***p<_.001

Table 7

Comparison of Correlation Obtained Using Original and Recoded Data in the Total Sample

Predictors	Total Sample		
	Recoded Subsample (n=81)	Original M-70 Scores (n=353)	With Failure M-70 Scores Added (n=374)
Tracking			
Tracking Composite	-.33**	-.25***	-.28***
One-Hand Tracking	-.30**	-.18***	-.21***
Two-Hand Tracking	-.33**	-.28***	-.29***
Target Shoot			
Time to Fire	-.30**	-.11	-.15**
Distance	-.14	-.10	-.11*
Cannon Shoot			
Absolute Mean Disc. Time	-.19	-.15**	-.16**
Spatial Test			
Spatial Reasoning	.25*	.14*	.15*
Object Rotation	.15	.17**	.21***
Orientation	.31***	.23***	.28***
Maze	.26*	.08	.13*
Map	.19	.16**	.20***
Assembling Objects	.11	.06	.10

*p <_ .05; **p <_ .01; ***p <_ .001
[a]Lowest score recoded to next lowest score.
[b]Scores for all failures recoded to 549.

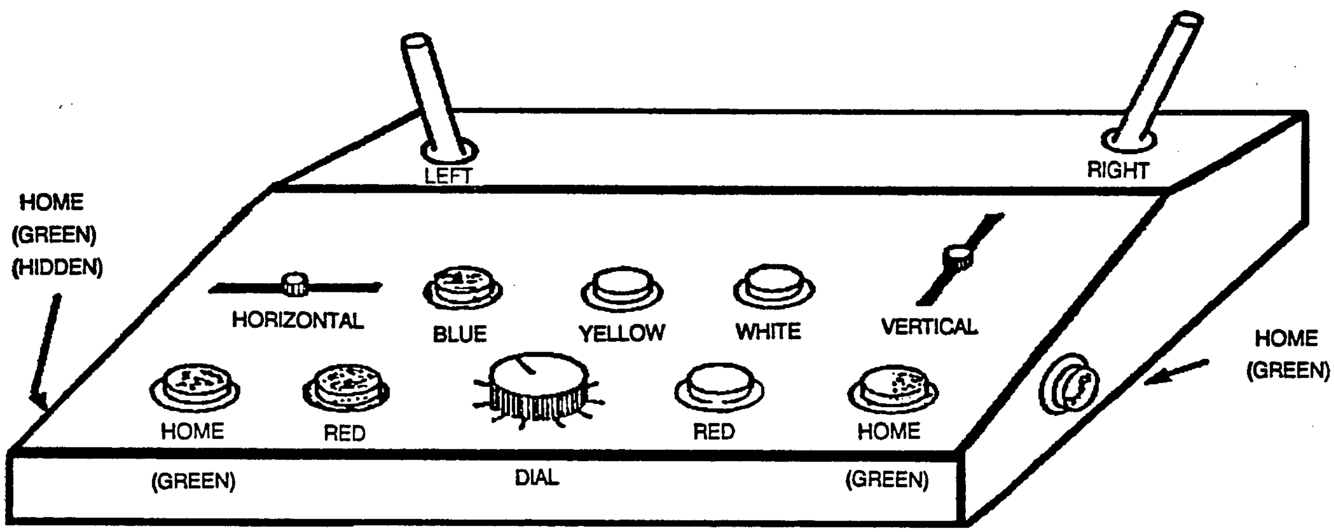


Figure 1. Response pedistal for computerized tests.

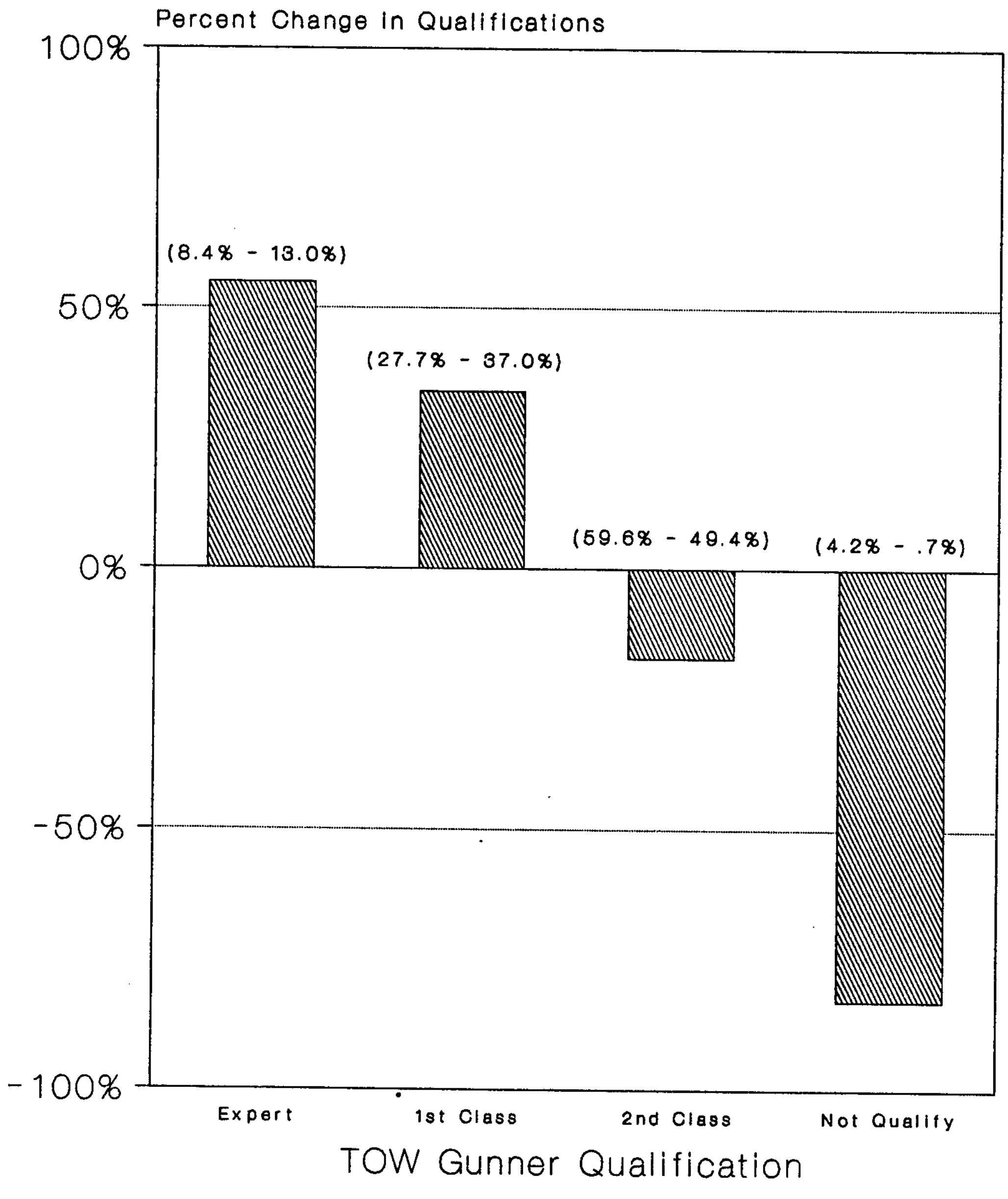


Figure 1. Expected increase in qualified TOW gunners because of psychomotor screening (L. Hanser, 1988).

Working Paper

RS-WP-87-Ø6

PROJECT A CONCURRENT VALIDATION RESULTS: RESPONSES TO CG TRADOC QUESTIONS

Frances C. Grafton, Darlene M. Olson, Newell K. Eaton, and
Lawrence M. Hanser

ARMY RESEARCH INSTITUTE

Jeffrey J. McHenry

AMERICAN INSTITUTES FOR RESEARCH

August 1987

REVIEWED BY:
MICHAEL G. RUMSEY
Research Psychologist
SCTA

APPROVED BY:
LAWRENCE M. HANSER
Chief, Selection and
Classification Tech Area

**U.S. Army Research Institute
for the Behavioral and Social Sciences**
5001 Eisenhower Avenue, Alexandria VA 22333

Introduction

The purpose of this paper is to provide a detailed response to questions about Project A raised by GEN Thurman during a briefing at Ft. Benjamin Harrison on 13 July 1987. GEN Thurman requested ARI to provide to him, by MOS, the complete data on the prediction of job performance using the ASVAB test. He expressed interest in seeing the relationships between AFQT category and MOS-specific task level performance for each of the Project A MOS.

A broad overview of Project A is provided in Appendix A, and copies of the slides from the 13 July 1987 briefing are contained in Appendix B. In addition to the specific analyses requested by GEN Thurman, we have also provided summaries of several analyses which provide for greater understanding of these data.

Background

The information reported in this paper is based on data collected during the concurrent validation (CV) phase of Project A. In the summer and fall of 1985 we tested approximately 9500 first tour soldiers, with a target sample size of 600 per MOS in our 19 MOS. The 19 MOS are listed on page 7 of Appendix B. These MOS were divided into two groups, with the primary distinction that hands-on tests were developed for the first group of nine MOS.

In addition to a four-hour battery of potentially useful new predictor tests to augment ASVAB in initial entry screening, we administered tests to soldiers in each of nine MOS in order to obtain three kinds of data on technical job performance. These were: (1) hands-on tests (HO) for approximately 15 tasks, (2) task-based job knowledge tests (K5) for these 15 tasks and an additional 15 tasks, and (3) multiple-choice training-based job knowledge items in a number of subtests (K3). In these nine MOS, performance testing took 12 hours over a day and a half. For these nine MOS, the average number of soldiers tested was 448. These are referred to as the Batch A MOS.

For soldiers in our remaining 10 MOS we collected the predictor tests and training-based job knowledge information only. In these MOS all testing took place in one day. These 10 MOS are referred to as the Batch Z MOS. All tests were coordinated with and approved by the MOS proponent before the data collection.

Primary Analyses

Relationships Between AFQT and Task Performance

In Appendix C are tables for soldiers in each AFQT category showing the mean percent correct or mean score on each task for each MOS in our sample. Task titles preceded by "HO" were tested hands-on while those preceded by "K5" were procedural knowledge paper-and-pencil tasks. We also provide the mean score for all 19 MOS on categories of the training-based subtests; these are designated with "K3" on the printouts.

We have marked each task or subtest with an "*" if there is an overall statistically significant ($p \leq .05$) increase in performance as a function of AFQT category. Within those tasks demonstrating an overall significant difference we have additionally marked the adjacent groups (II vs. IIIA, IIIA vs. IIIB, and IIIB vs. IV) showing a significant difference. Due to the generally small sample sizes in Category I, no significance tests were computed for Category I vs. Category II, and we did not include Category I soldiers in the overall significance tests. Of the 655 tasks overall, performance on 59% of the tasks increased with AFQT. Across the 19 MOS the lowest number of tasks showing a significant difference across AFQT categories was 33% for MOS 91A, and the highest was 93% for 54E MOS. Caution should be exercised in the interpretation of cells with few soldiers.

In addition to the individual tasks of each type of measure, several aggregate scores are also analyzed for each MOS. For all 19 MOS the training-based (K3) subtests are aggregated into three to six scores related to general job dimensions (Communications, Vehicle Maintenance/Operation, Basic Soldiering, Identify Target, Technical, and Safety). Overall MOS-specific scores for each type of task administered (HO, K5, and K3), total MOS-specific Core Technical Proficiency and General Soldiering Proficiency scores are analyzed whenever available. We were also able to merge these data with the SQT data file, and have shown for each MOS (except 54E and 91A) the mean SQT score for soldiers in each AFQT category.

In general, although some individual tasks may not show a difference across AFQT categories, the total scores for each type of task are significant for all MOS. The exception to this is the hands-on component for the nine Batch A MOS. The number of hands-on tasks showing significant differences among AFQT categories varies from a low of one out of fifteen for MOS 63B (seven percent) to a high of six out of seventeen (thirty-five percent) for MOS 13B and 95B. Five of the Batch A MOS (11B, 13B, 31C, 71L and 91A) show significant differences for the total Hands-On Performance score (HO Core Technical Proficiency) but

the other four (19E, 63B, 64C, and 95B) do not. As a basis for comparison, the number of job knowledge (K5) tasks showing significant differences varies from a low of 34% (MOS 31C) to a high of 93% (MOS 11B) and the number of training-based job knowledge (K3) subtests showing such differences varies from a low of 33% (MOS 51B) to a high of 100% (MOS 11B).

Additional Analyses

Because of the sheer volume and complexity of the task-level data, we have also provided several summary analyses to assist in understanding the full meaning of these very important data.

Interrelationships Among Performance Measures

The first of our additional analyses concerns our various measures of MOS performance. For each MOS, the scores from all of our performance measures are moderately to highly correlated. Because of these relations, more reliable measures of performance can be obtained by aggregating scores across tasks and measures. Samples of the results of these analyses are shown in Table 2 of the paper by Arabian and Mason, provided in Appendix C. The mean correlations are .68 for training-based knowledge (K3) with job knowledge (K5), .41 for training-based knowledge (K3) with hands-on (HO), and .48 for job knowledge (K5) with hands-on (HO). It should also be noted that SQT scores are moderately correlated with all three types of Project A performance measures.

Relationships Between AFQT and Core Technical Proficiency

In Appendix E are tables for each of the nine Batch A MOS. We completed these analyses by aggregating the performance measures to obtain an overall MOS-specific performance score (Core Technical Proficiency) for each soldier, and then divided the soldiers into three groups; those scoring in the upper, middle, and lower thirds, in each MOS, on the overall Core Technical Proficiency score. The specific meaning of the results of partitioning soldiers into thirds obviously is dependent on the quality of soldiers in the current force. That is, since today's Army has many high quality soldiers, those in the bottom third today are in reality much better performers than we would have seen several years ago. Nonetheless, we have found these distinctions useful in understanding the range of performance that exists even among high quality soldiers. If soldier quality were to be lowered, we would expect to see even larger differences in performance.

The tables show the relation between AFQT category and overall MOS-specific job performance (Core Technical Proficiency) for each of these MOS separately. They also indicate the number of soldiers in each AFQT category in each MOS sample. In almost

every case the percent of middle and high performing soldiers increases with increasing AFQT. We would place little importance on those few cases where this does not hold; they are cases where the sample size is too small to be reliable (e.g., two AFQT IVB soldiers in the MOS 13B table).

Expected Core Technical Proficiency by AFQT

Tables illustrating the results of regression analyses for the nine Batch A MOS are provided in Appendix F. These tables contain data on the expected overall Core Technical Proficiency for soldiers in each AFQT category having 10 or more soldiers. The predicted score is normalized with a mean of 100 and a standard deviation of 20. These findings show that across MOS the expected MOS-based technical performance of soldiers increases as a function of AFQT category.

Impact Analyses

The next set of analyses addresses the question of the impact of observed performance differences on the Army. These analyses are based on the partitioning of soldiers into thirds based on overall Core Technical Proficiency scores as described above.

What does it mean to have today's soldier in the top third or bottom third of a particular performance category in terms of scores on particular tasks? To help answer that question, we have provided a summary of these results for MOS 11B in Appendix D. This summary shows how soldiers in the upper, middle, and lower thirds, overall, performed on tasks/steps of the various Project A performance measures.

Relationships Between ABLE and Job Performance

The final set of analyses we present here concern a new test developed by Project A, the Assessment of Background and Life Experiences (ABLE). This test was developed to measure attributes that are related to the motivational, or "will do", aspects of job performance and is very valid for this purpose. Historically, ASVAB has not been a good predictor of attrition or other "will do" performance criteria. This is not surprising since ASVAB was designed to predict technical MOS performance, which it does very effectively, and not the motivational components of performance.

ABLE has three functional scales (1) Achievement Orientation, (2) Dependability, and (3) Physical Condition. Achievement Orientation predicts Effort and Leadership: A soldier's ability to exhibit effort across the entire range of job tasks, persevere under adverse or dangerous conditions, and provide peer support. Dependability predicts Personal Discipline: The degree to which a soldier exhibits self-control, adheres to Army regulations and

traditions, and avoids disciplinary actions. The Physical Condition score on ABLE predicts Physical Fitness and Military Bearing: The degree to which a soldier maintains good physical conditioning and displays proper military bearing.

Data on the relationship between scores on ABLE and the motivational components of performance are shown in Appendix H, pages 3-6. Note, for example, that soldiers scoring in the bottom 1/3 on ABLE have almost three times the Article 15/flag action rate of those in the top 1/3, while ASVAB does not differentiate between those with high and low disciplinary rates.

Data on the unique and incremental validities of ABLE are presented in Appendix H, page 7. When we examine ABLE as compared to education and ASVAB, we find that ABLE is the single strongest predictor of the three "will do" performance factors. It adds much more predictive power to ASVAB than education does.

Summary

From the data presented in this paper four conclusions emerge: (1) AFQT is closely related to technical proficiency, although this relationship varies with type of performance measure, as well as by task, (2) reasonable interrelationships have been found among Project A job performance measures; (3) the performance measures do a good job of differentiating performance among soldiers; and (4) a new Project A measure, ABLE, predicts the motivational aspects of job performance that the ASVAB was never intended to predict.

Information from the Project A concurrent validation data collection in 1985 discussed here, and from the follow-on major longitudinal validation data collection, scheduled for summer of FY88, will continue to provide a unique data set, unparalleled in richness for addressing recurring Army concerns. These data are invaluable to the Army in the areas of accession policy, setting standards for enlistment and reenlistment, predicting attrition and reenlistment, linking school training to field performance, and assisting in the accomplishment MANPRINT requirements.

OneStep™
Index System

Contents

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-17

Current Army Selection and Classification Research

Frances C. Grafton
Darlene M. Olson
Clinton B. Walker
Jane M. Arabian
Selection and Classification Technical Area

March 1989

REVIEWED BY:

Leonard A White

Leonard A. White
Research Psychologist
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

for Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Final
Sent 3/24/89

Introduction

Beginning in 1979 military testing has been the focus of increasing attention, from both the Department of Defense and Congress. The reasons for this interest are varied and include:

1) The discovery in 1979 that the then operational selection test battery for military enlistment was misnormed.

2) The switch in reference groups for military tests from the 1944 mobilization population (all men, including officers, serving under arms during that year) to the 1980 youth population (all 17- to 23-year olds in 1980 including both males and females).

3) A Congressional mandate to link entry test scores directly to on-the-job performance rather than to traditional training performance.

4) Intense interest in combining computer technology with emerging developments in testing theory.

As an integrated response to the various concerns, the U.S. Army Research Institute designed Project A, "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel," a multi-year, multi-million dollar research effort. Project A had many objectives including (1) continued validation of the operational Armed Forces Vocational Aptitude Battery (ASVAB), (2) development and validation of new cognitive and non-cognitive measures that might expand the coverage of the predictor domain beyond that of the ASVAB in predicting total soldier performance, and (3) development of a comprehensive set of performance criteria including paper-and-pencil tests of school knowledge and job knowledge, hands-on tests of job performance, and behaviorally-anchored rating scales.

Project A was initiated in 1982 and will be completed in 1989. While Project A focused primarily on Army enlisted personnel in their first tour of service, a second research program, "Building and Retaining the Career Force," is scheduled to begin in 1990. This effort will focus on the Army's career force, those soldiers who reenlist and remain beyond one tour of duty. Soldiers who comprise the career force provide not only leadership and continuity to the Army as an institution, but are crucial in insuring that the Army meets its battlefield mission.

Another accomplishment by the Army in military testing during the last decade has been the development and implementation of the Computer Adaptive Screening Test (CAST). CAST is used in recruiting stations to determine whether a prospect is likely to score well enough on the ASVAB to be considered for enlistment. Prior to implementation of CAST all prospects' eligibility to enlist had to be determined by transporting them to a testing

site to take the three-and-a-half hour ASVAB. The use of CAST has reduced the number of ASVAB tests given to Army prospects from over one million per year in the early 1980s to approximately 400,000 per year currently, while the number of accessions per year has remained about the same.

Recognizing that large scale efforts like Project A will not be feasible for all Army jobs, an exploratory research effort, "The Army Synthetic Validation Project," is currently underway. This program is designed to (1) develop procedures for deriving prediction equations for MOS using primarily the extensive data gathered on the relatively small sample of MOS in Project A, and (2) develop procedures for setting selection standards linked to job performance.

The remainder of this paper will more fully describe these efforts as well as some of the resulting products that have been implemented either operationally or in testbeds. Future plans in the area of military testing will also be discussed.

Project A

In Project A, the research objectives are to:

1) Develop new, comprehensive measures to cover the job performance domain in the Army. These measures include both Army-wide and MOS-specific rating scales, written school and job knowledge tests, and direct hands-on measures of MOS-specific task proficiency.

2) Validate ASVAB against both existing and new performance measures.

3) Develop and validate new selection and classification measures against existing and new performance measures.

4) Validate intermediate criteria (such as training performance), as predictors of later criteria, such as first-tour job performance, so that better informed decisions on reenlistment can be made.

5) Examine the validity and utility of alternative procedures for making operational selection and classification decisions in the Army.

The Project A research design consists of three main stages of data collection and analysis in an iterative progression of development, testing, evaluation, and further refinement of the selection/classification instruments (predictors) and measures of job performance (criteria). In the first stage, individual-level data from FY 81-82 were examined to explore relationships between soldiers' ASVAB scores and their later performance in training and on first-tour Skill Qualification Tests (SQT).

Prior to the second stage of the data collection, 19 Military Occupational Specialties (MOS) were selected as a representative sample of the Army's 250+ entry-level jobs. These MOS were selected based ~~on 1) a~~ clustering of MOS based on similarity of job content job content, 2) the fact that these MOS are representative ~~of Army~~ jobs in general and account for over 45% of the Army's accessions, and 3) the recommendations of General Officers representing major Army Commands at that time. Nine of the 19 MOS were then selected to have both paper-and-pencil and hand-on tests of job performance in addition to the school knowledge tests and ratings administered in all of the MOS. The nine fully-tested MOS in Project A are: Infantryman (11B), Cannon Crewman (13B), Tank Crewman (19E), Radio Operator (31C), Light Wheel Vehicle Mechanic (63B), Administrative Specialist (71L), Motor Transport Operator (88M), Medical Care Specialist (91A), and Military Police (95B).

The second stage of the Project A research, the Concurrent Validation (CV), was conducted in FY 1985. During the CV, over 9400 soldiers in the 19 MOS were administered the new Project A predictor tests, which included measures of spatial abilities, temperament and vocational interest, as well as computerized tests of perceptual and psychomotor skills. These tests were designed to expand the predictor domain in terms of individual characteristics and attributes that might be important for selection in predicting more aspects of performance. The new predictors were intended to supplement those cognitive abilities already assessed by ASVAB.

Concurrently, the CV soldiers, who had been in the Army 18 to 27 months, were administered a comprehensive set of job performance measures including supervisory and peer ratings,written school and job knowledge tests, and MOS-specific hands-on task proficiency measures.

One of the major scientific contributions of the Project A research to date is a comprehensive modeling of the job performance domain. The criterion development efforts in the project were driven by the idea that job performance is multi-dimensional and that performance is best measured through a variety of methods. Analyses of the criterion data collected during the CV resulted in five empirically derived dimensions that represent overall Army job performance. The five Project A job performance dimensions are:

1) MOS-Technical Knowledge and Skill. The proficiency with which a soldier performs the technical tasks that are critical and central to his/her MOS.

2) General Soldiering Proficiency: How well a soldier executes common soldiering tasks such as first aid and land navigation.

3) Effort and Leadership: The degree to which a soldier

exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers.

4) Personal Discipline: The degree to which a soldier adheres to Army regulations, exhibits self-control and does not create disciplinary problems.

5) Physical Fitness and Military Bearing: The degree to which a soldier maintains an appropriate military appearance and remains in good physical condition.

The first two performance dimensions involve a soldier's ability to perform the technical requirements of his/her job and are often referred to as "can do" factors. The remaining three performance dimensions are more attitudinal or motivational in nature and are often called "will do" factors.

The third stage of the Project A research, known as the Longitudinal Validation (LV), started in 1986 with the administration of the battery of new Project A predictor tests to approximately 55,000 new recruits in 21 MOS. At the end of his/her Advanced Individual Training (AIT) or One Station Unit Training (OSUT), each soldier was administered the school knowledge test for his/her MOS, and a set of rating scales was collected from supervisors and peers. Approximately 11,000 of these soldiers were followed into their first-tour field assignment and have been administered a set of MOS-specific and Army-wide job performance measures during FY 88 and FY 89. In addition, about 1,000 of the soldiers from the CV sample who reenlisted have been assessed on a battery of performance measures appropriate for second-tour including rating scales, written job knowledge tests, hands-on measures, and leadership measures.

Results and Products

During the execution of the research plan, Project A made many scientific contributions to the field of Industrial and Organizational psychology and so far has delivered four major products to the Army.

ASVAB. One product which resulted from the Concurrent Validation involves improvements in the computation and use of ASVAB Aptitude Area (AA) Composites. These composites are used to determine eligibility for training in Army jobs. Beginning with the operational implementation of ASVAB 11/12/13/14 in October 1984 the formulae for the Clerical (CL) and the Surveillance/Communications (SC) composites were changed. These changes resulted in both better accuracy and improved fairness in the prediction of job performance for minorities in several MOS. In addition, a change in the computation of the Mechanical Maintenance (MM) composite and recommended changes in required aptitude areas for approximately 50 MOS are scheduled for

implementation in the near future. Annual savings from the changes in composites have been estimated at $25 million.

The Enlisted Personnel Allocation System (EPAS) is a new assignment system that will more efficiently match qualified Army applicants to jobs for which they are best qualified, maximizing performance and minimizing attrition. Thus, with significant improvements in classification composites (from Project A) and in assignment (from EPAS) the Army can put the "right" person in the "right" job at the "right" time. The potential savings to the Army from using ASVAB (including the improved composites) and the Enlisted Personnel Allocation System to fill accession requirements is estimated to be more than $480 million annually.

Results from the Project A CV have demonstrated conclusively that ASVAB is an excellent test of the "can do" or more technical task-based requirements of Army jobs. The mean validity coefficient between scores on ASVAB and Core Technical performance is .57 across the 19 MOS in the Concurrent Validation. Since ASVAB was designed as a general cognitive test, it is not expected to be a good measure of more specialized spatial abilities, psychomotor skills, motivation, interests or leadership. As a means of supplementing ASVAB, the new Project A tests measure these characteristics very well.

Spatial, Psychomotor and Perceptual Tests. Currently, ARI is supporting implementation of Project A's spatial and computerized psychomotor and perceptual tests in USAREUR and at training posts in CONUS.

These implementations were inspired by ARI selection research on tank and anti-tank gunners. In 1986, Project A's new tests had a multiple R of .76 against accuracy on high-tech simulators of tank gunnery for 95 Armor officers at Fort Knox. Cognitive ability, which was tested as well, did not contribute to the prediction. In 1987, 300+ new recruits at Fort Benning took the same battery before training in anti-tank gunnery. Several of the new tests strongly predicted accuracy in firing the TOW (Tube-launched, Optically-tracked, Wire-guided) missile simulator.

Based on these results, in December 87, CG TRADOC ordered implementation of Project A tests at four posts that train on high-tech weapons systems. ARI installed 1- and 2-hand tracking, maze, and mental rotation tests at Forts Knox, Benning, and Bliss in February 88.

At the Infantry and Armor sites, the earlier results have been confirmed. For 1,065 TOW students to date, 41% of those passing a cut score on the predictors qualified as gunners in the minimum time, as opposed to 24% of those not tested on the psychomotor/spatial predictors and 12% of those scoring below the cut score. Those scoring above the cut score qualified higher as well as faster: 48% attained the upper two levels of

accuracy as contrasted with 36% of those not tested and 31% of those below the cut score. In terms of validities, accuracy was predicted at .37 by the Project A tests, .29 by ASVAB GT, and .38 by the two together. In mid-88, these training data were confirmed in live fire at Fort Benning: for 60 students firing one live TOW each at a moving target at 6,000', P(hit) was .85 for those who met the cut on the Project A tests and .73 for those who did not.

At Fort Knox, Armor recruits (N = 500+) scoring in the upper third on the new predictors early in training later had gunnery hit rates 16% higher than those in the lower third, exactly repeating the results in the 1986 research. The new predictors correlated .54 with speed/accuracy, compared with a .34 validity for the ASVAB GT composite. Combined, the validity was .55, reconfirming the importance of spatial and psychomotor skills in predicting gunnery performance. At both posts, the new results were so positive that the test scores were incorporated into the decision-making process.

Validities at Fort Bliss were not significant, and testing there has been suspended. In contrast to tank and anti-tank weapons, which rely heavily on tracking, the air defense systems tested use fire-and-forget weapons against fleeting targets. Thus these systems rely more on skills like vigilance and target identification.

Completing the implementation requests from CG TRADOC, the Field Artillery School, Fort Sill, is starting to administer a broader battery of the new tests to recruits in meteorology, surveying, radar range-finding, and artillery spotting (MOS). Validation will be against performance at the end of training.

Currently, the active forces are transitioning to the Bradley Fighting Vehicle. In USAREUR, V Corps has elected to use a battery of spatial and psychomotor tests to inform their selection of Bradley gunners in the 3rd Armor Division and the 8th Infantry Division.

A need for improved selection has been found also at the Special Forces School, where attrition, due primarily to failures in land navigation, is costly. In a new pilot project, ARI has installed three spatial tests at Fort Bragg to identify good land navigators.

In review, early positive results are being replicated in the implementations; spatial/psychomotor abilities strongly predict differences in gunnery performance; and utilization is spreading. Budget reductions, however, threaten the survival of the implementation in the training base. In the future, the tests could have the greatest impact if they were administered before enlistment. In that case the Army Reserve and National Guard could be served as well, and initial person-job matching could be strengthened in many MOS by adding spatial and/or

psychomotor abilities to the profile of aptitudes.

Assessment of Background and Life Experiences (ABLE). The ABLE is a 30-minute, multiple choice, non-cognitive test designed to measure temperament, personal history, and adaptability. In Project A, ABLE was shown to improve significantly the prediction of the motivational components of performance. Scores on the Adjustment scale of ABLE are strongly related to 12-month attrition. Recruits who have very low Adjustment scores have attrition rates that are two to three times higher than soldiers with high Adjustment scores.

The Dependability scale of ABLE predicts in-service disciplinary problems. For example, soldiers having low Dependability scores receive significantly more Articles 15 than those with high scores. Conversely soldiers with high Dependability scores are more often viewed as having potential of becoming high performing NCOs.

In sum, the ABLE shows promise to augment the Services' capability to identify the most qualified applicants in all AFQT categories and educational levels by predicting the "will do" components of performance. Based on these encouraging results, the ABLE is included in a joint Service adaptability screening instrument that will become operational in 1990.

The Longitudinal Research Data Base. The fourth and perhaps most enduring product from the Project A research effort is the Longitudinal Research Data Base (LRDB). The LRDB is a permanent storehouse of empirical information on Project A, unparalleled in its richness for addressing recurring Army concerns. These data are invaluable to the Army in the areas of accession policy, setting standards for enlistment and reenlistment, predicting attrition, linking school training to field performance, and linking characteristics at entry into the military to performance in first and second tours of duty.

Building and Retaining the Career Force

At the end of the Project A Longitudinal Validation data collections in FY 89, the first phase of the Army's research program to build a new selection and classification system ,based on job performance, will be completed. The second phase, which will include measures of critical tasks for soldiers in their second tour, will be initiated in FY 90.

The research objectives planned for Building and Retaining the Career Force are to:

1) Develop a set of measures for selecting and classifying enlisted personnel in order to optimize second tour soldier performance without sacrificing first tour performance.

2) Analyze longitudinal data from Project A and other sources to develop information, procedures, and recommendations for implementing new selection and classification measures in building the Army's career force.

3) Determine the validity of ASVAB, new predictors, training performance, and first tour performance in predicting future performance (including second tour).

Results generated from both Project A and Building and Retaining the Career Force will be useful to Army policy makers in setting enlistment standards and to Army training proponents in setting quality distribution goals during years of a declining manpower pool. The outcomes will also be helpful to DoD in providing (1) more tools for conducting selection and classification testing, and (2) scientific data to use in responding to Congress and the general public on questions concerning military testing.

Computerized Adaptive Screening Test (CAST)

The Computerized Adaptive Screening Test (CAST) is a tool for recruiters to estimate the likely performance of a prospect on the Armed Forces Qualification Test (AFQT) of the ASVAB. As the name suggests, CAST is a computer adaptive test, and it is administered in recruiting stations. The computer selects succeeding items based on responses to previous items. Because item difficulty is matched to examinee ability, maximum information is obtained from each item. Thus, a short CAST can provide as much information as a longer paper-and-pencil test.

CAST was preceded by the Enlistment Screening Test (EST). The EST was a good predictor of individuals' scores on the AFQT, but it had two problems. First, recruiters felt it was too long, and second, the EST had to be hand scored. The implementation of JOIN (Joint Optical Information Network) provided computers for recruiting stations, and computer-administered testing became feasible. Whereas the EST, a paper-and-pencil test, takes 45 minutes, CAST can be completed in 15 minutes.

CAST was initially developed by the Naval Personnel Research and Development Center (NPRDC) in a cooperative arrangement with ARI. A field test on 364 Army applicants showed a validity of .85 using AFQT as the criterion. This was an encouraging finding, comparable to the validity of .83 for EST. Two later cross-validations on larger samples yielded comparable results (i.e. validities of .80 and .79).

The initial implementation of CAST received a generally positive reception from recruiters and was successful in predicting operational AFQT. However, some improvements were necessary, leading to recent refinements in CAST. The number of items in CAST was increased from approximately 300 to over 500. This increase was needed to reduce the likelihood that any

particular item would be compromised as a result of overuse. In addition, a better format for reporting results was needed. The original display predicted a specific AFQT score. What the recruiter and prospect really wanted to know, however, was the likelihood that the prospect would score in certain critical ranges of the AFQT. Accordingly, the final display was redesigned to meet that need. The current version of CAST also has been calibrated to the new AFQT introduced operationally in January 1989.

Synthetic Validation Project

The Army's Synthetic Validation (SynVal) Project is an exploratory research effort currently underway. The concept of synthetic validity was first introduced by Lawshe (1952) as a middle ground between situational validity, which requires costly validity analyses (empirical validation) for each job, and generalized validity, which assumes the validity of a test across similar jobs. The Project A research, with its comprehensive set of new predictor tests (non-cognitive, spatial, and psychomotor) in addition to the existing ASVAB and extensive job performance measures, afforded a unique opportunity to examine SynVal as a cost-effective alternative to empirical validity studies for a large number of MOS.

This project has two main goals:

1) Develop procedures for identifying job performance prediction equations for new MOS, low-fill MOS, and for other MOS when it is impractical or too costly to derive empirical prediction equations.

2) Develop procedures for establishing cut scores on predictor tests that are linked to job performance and that identify both minimally qualified and highly qualified recruits.

The research design has three phases, each with a data collection. The initial phase concentrates on MOS in the Project A sample that had the full complement of criterion measures; the second phase adds MOS which had reduced criterion testing; and the final phase will apply the SynVal procedures to at least one MOS that was not included in Project A. Comparisons will be made among job component models and among standard setting procedures.

Three job component models have been developed and tested: (1) Job Behaviors Model, (2) Attribute Model, and (3) Job Tasks Model. The results of phase one have been extremely encouraging. Army subject matter experts (SMEs) were able to reliably use the three models to identify job components (in terms of importance or estimated validity) for each of the three MOS. The resulting job description (component) profiles differed systematically across MOS. Furthermore, the average "synthesized" validity

(across the three phase one jobs and models) compared favorably with the average empirical validity (.50 vs .67, respectively). However, given the exploratory nature of these findings, it would be premature to exclude any model or weighting scheme.

With respect to the second goal, an extensive literature review found little research on procedures for setting cut scores that are linked to job performance. The project assumes that requirements for different levels of job performance must be determined first. Then cut scores could be set on the predictor (selection) measure(s) to optimize the likelihood of obtaining individuals who would meet the specific performance requirements.

Four performance levels have been objectively defined: Unacceptable, Marginal, Acceptable, and Outstanding. The three standard setting approaches developed and tested in phase one were: (1) Soldier-based, (2) Task-based, and (3) Critical Incident-Based. Although mixed results have been obtained, the existing literature on standard setting also consistently indicates differences in obtained standards with different techniques. The critical-incident approach resulted in the most lenient standards for minimal performance followed by the soldier-based approach with the task-based approach being the most stringent. However, the task-based approach resulted in more variation (less agreement) among the judges while the critical-incident approach showed the greatest agreement.

A third of the way through this project, the results are encouraging. In making use of the Project A LRDB to extend validity information from a relatively small sample to potentially all entry-level MOS, SynVal represents a major return on the Project A investment. Future research, building on this foundation, might include testing the transferability of the SynVal methodologies to the other Services as well as to the weapon system acquisition process.

Working Paper

RS-WP-84-10

ASVAB RETEST SCORE GAINS

Karen J. Mitchell & Lawrence M. Hanser
SELECTION AND CLASSIFICATION TECHNICAL AREA

March 1984

Francis J. Medland

Reviewed by
Frances C. Grafton
Plans, Programs
and Operations

Approved by
for Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

V

PO 9628

ASVAB RETEST SCORE GAINS

Karen J. Mitchell
Lawrence M. Hanser

INTRODUCTION

The Armed Services Vocational Aptitude Battery (ASVAB) was introduced in January 1976 as the single Department of Defense selection and classification battery. It is administered to over one million applicants for military service each year. In FY83 over 450,000 nonprior-service Army applicants were tested on ASVAB, forms 8, 9 and 10.

The ASVAB is composed of eight power and two speed tests. Four ASVAB subtests are combined to form the Armed Forces Qualification Test (AFQT), a general measure of trainability and the primary criterion for enlistment. The AFQT cutoff for FY83 Army accession was set at the 16th percentile for high school graduates and the 31st percentile for nongraduates. In FY83 85.1 percent of the Army applicants met AFQT standards on initial testing while 13.4 percent failed to meet the AFQT enlistment criterion on the initial score-of-record. The remaining 1.4 percent were selected for verification retesting. Examinees are asked to verify test scores by reexamining on the AFQT when inconsistencies in performance are noted on various parts of the test.

During FY83 11.8 percent of Army applicants took the ASVAB two or more times. Examinees may retest because (1) they fail to meet the enlistment cutoff on initial testing, (2) they wish to increase their scores on the aptitude area composites which are used for occupational classification, or (3) they are asked to repeat the test for verification purposes. Of the approximately 53,000 Army applicants who retested in FY83, 26.6 percent had failed on initial testing; 62.1 percent had passed, and 11.4 percent were verification retesters.

Sims and Maier (1983) and Friedman (1983) have examined test score gains for examinees taking ASVAB two or more times. Their work shows that applicants show consistent and meaningful score gains on retesting. It is hypothesized that examinees become familiar with test booklet formats, answer sheet layouts, test content, pacing requirements, etc. on multiple testings. The examinee, thereby, acquires test-taking skills and content knowledge that facilitate performance on subsequent testings. These test score gains are expected to be approximately equivalent across the subtests of ASVAB.

Sims and Maier (1983) and Friedman (1983), however, have shown that retest gains for the two ASVAB speed subtests, Numerical Operations (NO) and Coding Speed (CS), exceed those for the ASVAB power subtests and that the speeded test gain (in sigma units) is roughly twice the size of the gain observed for the other tests. The purpose of this research is to examine the impact of these gains on enlistment qualification in the Army.

METHOD

Data for FY83 nonprior-service Regular Army applicants who tested two or more times on ASVAB 8/9/10 were examined. AFQT scores were computed at initial testing and at retest using the operational procedure. AFQT indices were also computed using the retest Arithmetic Reasoning (AR), retest Word Knowledge (WK), retest Paragraph Comprehension (PC), and initial NO scores. This computing scheme was developed to minimize the disproportionate speeded test score gain reported by Sims and Maier (1983) and Friedman (1983).

RESULTS AND DISCUSSION

Mean AFQT scores computed at initial testing and retest using the operational procedure appear at Table 1 for retesters who (1) failed to meet accession standards, (2) met the AFQT criterion, and (3) were asked to verification test at initial testing. Mean AFQT gains from initial to retest also appear for the three groups.

TABLE 1
MEAN AFQT INITIAL AND RETEST PERCENTILE SCORES

	Initial	Retest	Mean Difference
Failed	15.81	22.89	7.07
Passed	31.72	37.97	6.25
Verified	39.18	45.76	6.58

Of the retesters who failed to meet the AFQT cutoff on initial testing, 45.6 percent qualified on their retests. Of these applicants 43 percent enlisted.

Table 2 presents AFQT data for initial testing and for a second testing with AFQT computed using the retest AR, WK, and PC scores and initial NO score. Mean AFQT scores and mean gains from initial to second testing are displayed.

TABLE 2
MEAN AFQT INITIAL AND RECOMPUTED RETEST PERCENTILE SCORES

	Initial	Retest	Mean Difference
Failed	15.81	20.22	4.41
Passed	31.72	35.61	3.90
Verified	39.18	42.69	3.51

The mean retest gain reported in Table 2 is from 2.3 to 3.1 points lower than that presented in Table 1. Of the retesters who had failed to meet enlistment standards initially, 35.6 percent qualified for enlistment when AFQT was computed in this way. Operationally, this scoring procedure may be implemented by maintaining initial NO and CS scores in an applicant's retest record and using the first-time speed scores and retest power scores in computation of AFQT and the occupational composites.

REFERENCES

Friedman, D., Streicher, A., Wing, H., Grafton, F. C. & Mitchell, K. J. (1983). Study of the Reliability of Scores for Fiscal Year 1981 Army Applicants: Armed Services Vocational Aptitude Battery, Forms 8, 9, and 10. Alexandria, Virginia: Army Research Institute, Draft Technical Report.

Sims, W. H. & Maier, M. H. (1983). The Appropriateness for Military Applications of the ASVAB Subtests and Score Scale in the new 1980 Reference Population. Alexandria, Virginia: Center for Naval Analyses, Memorandum 83-3102.

Working Paper

RS-WP-84-07

THE 1980 YOUTH POPULATION NORMS: ENLISTMENT AND OCCUPATIONAL CLASSIFICATION STANDARDS IN THE ARMY

Karen J. Mitchell and Lawrence M. Hanser
SELECTION AND CLASSIFICATION TECHNICAL AREA

February 1984

Reviewed by
Frances C. Grafton
Personnel Policy Research Group

Approved by
for Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

and aptitude area composites (Maier & Sims, 1982; Ree, Valentine & Earles, 1983).

Early in 1983 Sims and Maier reported apparent score pattern discrepancies on the speeded subtests between the male NORC youth sample and a series of male military applicants. The discrepancy resulted in lower scores for the 1980 sample males on the two speeded subtests when compared to a sample of male military examinees. An analysis of possible sources of the discrepancy revealed that there were format differences between answer sheets used in the NORC field test and operational answer sheets used for military entrance testing.

In August of 1983, Earles, Guiliano, Ree & Valentine outlined a plan for estimating the differences in performance on the speeded subtests attributable to the non-standard NORC answer sheet. A study was conducted in the fall of 1983 at the Military Enlistment Processing Stations to develop corrections for the NORC speeded subtests and to adjust the percentile conversion tables. The corrected subtest and composite norms became available to the Service Labs in January of 1984. They will be implemented in October 1984 with the introduction of ASVAB forms 11/12/13.

PROBLEM

Since 1944 various forms of AFQT and ASVAB have been calibrated to the World War II mobilization population. The use of a constant reference population allows the Services to (1) track the quality of applicants and recruits over time, and (2) maintain enlistment and occupational classification standards across test forms and over time. With the introduction of a new normative base for ASVAB, research is needed to document the comparability of applicant and accession quality based on the World War II and NORC reference groups. Data are also needed on the appropriateness of maintaining current enlistment and occupational classification standards after October 1984.

METHOD

Scores of the NORC sample on the power subtests were weighted and converted to subtest standard scores using both the World War II and NORC norms. The NORC sample raw scores on the speeded subtests were converted to equated raw scores based on the January 1984 tables and then converted to subtest standard scores using both sets of norms. This procedure yielded two sets of subtest standard scores for the NORC sample: (1) standard scores referenced to the WWII population and (2) standard scores referenced to the 1980 Youth Population. Mean subtest standard scores were computed for these two sets of scores.

The subtest standard scores were then combined to form the ten Army aptitude area composites. These standard score sums were converted to Army

standard scores using the World War II and NORC norms. Mean composite scores were computed for each. Subtest raw scores for Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Numerical Operations (NO) subtests were also summed and the sum converted to percentiles for the AFQT index. Mean AFQT scores were computed using both norms.

The composite scores were computed and examined for each of six sex/ethnicity subgroups: (1) white males, (2) black males, (3) Hispanic males, (4) white females, (5) black females, and (6) Hispanic females. Finally, the scores at each of the current occupational classification cutpoints for the ten aptitude areas were examined in order to compare the differences between scoring tables based on the two different norm groups. To the extent that the ability scores assigned by the new norms differed from the ability scores assigned on the basis of the historical norm group, questions of adjusting the current enlistment and occupational classification standards would arise.

RESULTS AND DISCUSSION

Mean subtest standard scores using the NORC and World War II norms appear at Table 1. This information is reported graphically at Figure 1. The mean NORC-assigned verbal, math, and speeded subtest scores were lower than the World War II-converted means. This means, for example, that an examinee receiving a Numerical Operations (NO) score of 50 using the new norms would have received a 51 in terms of the World War II subtest norms. An examinee receiving a 50 on Mathematics Knowledge (MK) using the new norms would have received a 52 using the historical conversions.

The converse is true for the technical subtests, Auto-Shop Information (AS), Mechanical Comprehension (MC), and Electronics Information (EI). The assigned NORC score was from two to four points higher than the World War II-assigned score. An applicant receiving a score of 50 on AS would have received a score of 46 using the World War II norms. The mean General Science (GS) scores using the two conversions were roughly equivalent. Mean aptitude area composite scores appear at Table 2. The data are presented graphically at Figure 2. The mean NORC-assigned composites for Clerical (CL), General Technical (GT), and AFQT were lower than the traditional norm scores. Thus, an applicant receiving a score at the 51st percentile on AFQT using the 1980 norms would be at the 52nd percentile on the WWII norms. For the remainder of the composites, the NORC-assigned composite score exceeded the World War II-converted score. The mean aptitude area differences ranged from one half of an Army Standard Score point on Field Artillary (FA) to six standard score points on Mechanical Maintenance (MM).

Table 3 presents the mean aptitude area score differences for the sex/race subgroups. The entries are the differences between the NORC- and World War II-assigned aptitude area scores. In all cases the historical

score is the subtrahend. The differences are presented pictorially at Figure 3. Examination of differences across the rows revealed no apparent subgroup patterns.

Table 4 contains the NORC aptitude area cutoff scores that would have to be implemented to maintain ability levels equivalent to the enlistment and occupational classification standards that are currently in place. Table 5 provides the same data with the NORC aptitude area cutoff scores rounded to five-point score intervals.

REFERENCES

Earles, James A., Giuliano, Toni 1LT, Ree, Malcolm J. & Valentine, Lonnie D. (1983). The 1980 Youth Population: An Investigation of Speeded Subtests. Brooks, AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Kass, Richard A., Mitchell, Karen J., Grafton, Frances C. & Wing, Hilda (1982). Factor Structure of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10: 1981 Army Applicant Sample. Alexandria, Virginia: Army Research Institute, Technical Report 581.

Maier, M. H. & Sims, W. H. (1982). Constructing an ASVAB Score Scale in the 1980 Reference Population. Alexandria, Virginia: Center for Naval Analysis, Memorandum 82-3188.

Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery (1982). Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Ree, M. J., Valentine, L. D., Jr., & Earles, J. A. (1983). The 1980 Youth Population: A Verification Report. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Sims, W. H. & Maier, M. H. (1983). The Appropriateness for Military Applications of the ASVAB Subtests and Score Scale in the new 1980 Reference Population. Alexandria, Virginia: Center for Naval Analyses, Memorandum 83-3102.

Table 1

MEAN SUBTEST STANDARD SCORES

NORC and WWII Norms

Subtest	NORC	WWII
GS	49.90	49.63
AR	50.01	50.33
WK	49.98	50.81
PC	50.04	51.47
NO	50.04	51.14
CS	49.98	52.89
AS	49.97	46.26
MK	49.96	51.84
MC	49.91	47.55
EI	49.95	47.98

Table 2

MEAN APTITUDE AREA COMPOSITE SCORES

NORC AND WWII NORMS

AA	NORC	WWII	SUBTESTS
CO	99.99	96.4	AR, AS, MC, CS
FA	99.99	99.58	AR, MK, MC, CS
EL	100.0	98.07	AR, EI, MK, GS
OF	100.0	95.53	NO, VE, MC, AS
SC	100.04	98.37	NO, CS, VE, AS
MM	100.02	93.77	NO, EI, MC, AS
GM	100.0	95.81	MK, EI, GS, AS
CL	100.0	101.23	NO, CS, VE
ST	100.0	98.39	VE, MK, MC, GS
GT	100.0	100.19	VE, AR
AFQT	50.65	52.13	VE, AR, NO

Table 3

MEAN APTITUDE AREA DIFFERENCES

BY RACE/SEX SUBGROUPS

AA	WHITE MALES	BLACK MALES	HISPANIC MALES	WHITE FEMALES	BLACK FEMALES	HISPANIC FEMALES
CO	3.8*	3.57	3.64	3.43	3.27	3.2
FA	.48	1.55	1.06	-.10	1.07	.85
EL	1.71	3.34	2.95	1.43	3.58	3.43
OF	4.23	3.73	4.14	5.1	3.52	3.67
SC	1.57	1.04	1.43	1.95	1.47	1.41
MM	6.13	6.04	6.18	6.52	5.65	5.67
GM	4.12	4.82	4.51	3.97	4.86	4.74
CL	-1.09	-.51	-.50	-1.8	-.31	-.64
ST	1.53	2.43	2.14	1.35	2.26	2.12
GT	-.58	.81	.56	-.23	.72	.63
AFQT	-1.28	-1.86	-1.79	-1.49	-1.86	-1.86

*NORC - WWII

Table 4

Equivalence of NORC and WWII Aptitude Area Score Cutoffs

WWII Cut Off		80	85	90	95	100	105	110	115	120
NORC CUTOFF	AA CO	84	88	93	98	103	110	114	119	123
	FA	81	85	89	94	99	104	109	115	121
	EL	83	87	90	96	100	105	110	116	122
	OF	85	90	95	101	107	111	115	119	123
	SC	85	90	94	98	103	108	112	115	120
	MM	87	92	96	102	108	113	117	121	125
	GM	84	89	94	99	104	110	114	118	124
	CL	83	88	91	96	101	104	108	112	116
	ST	82	87	91	96	101	107	111	116	122
	GT	81	86	90	96	101	105	110	115	120

Table 5

Equivalence of Rounded NORC and WWII Aptitude Area Score Cutoffs

WWII Cut Off		80	85	90	95	100	105	110	115	120
	AA/CO	85	90	95	100	105	110	115	120	125
	FA	80	85	90	95	100	105	110	115	120
	EL	85	85	90	95	100	105	110	115	120
	OF	85	90	95	100	105	110	115	120	125
NORC CUTOFF	SC	85	90	95	100	105	110	110	115	120
	MM	85	90	95	100	110	115	115	120	125
	GM	85	90	95	100	105	110	115	120	125
	CL	85	90	90	95	100	105	110	110	115
	ST	80	85	90	95	100	105	110	115	120
	GT	80	85	90	95	100	105	110	115	120

Figure 1
MEAN SUBTEST SCORES
NORC AND WWII NORMS

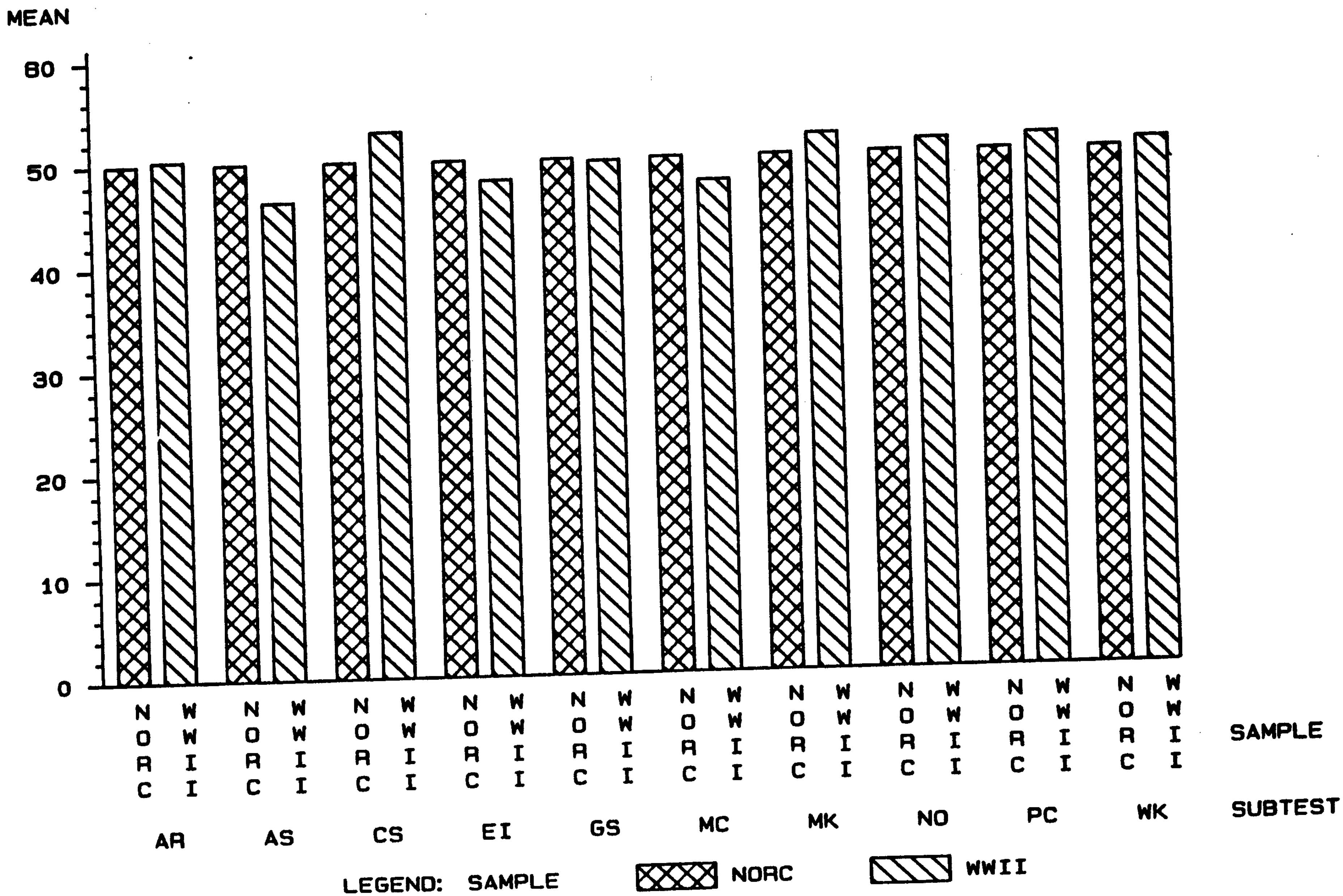


Figure 2
MEAN APTITUDE AREA SCORES
NORC AND WWII NORMS

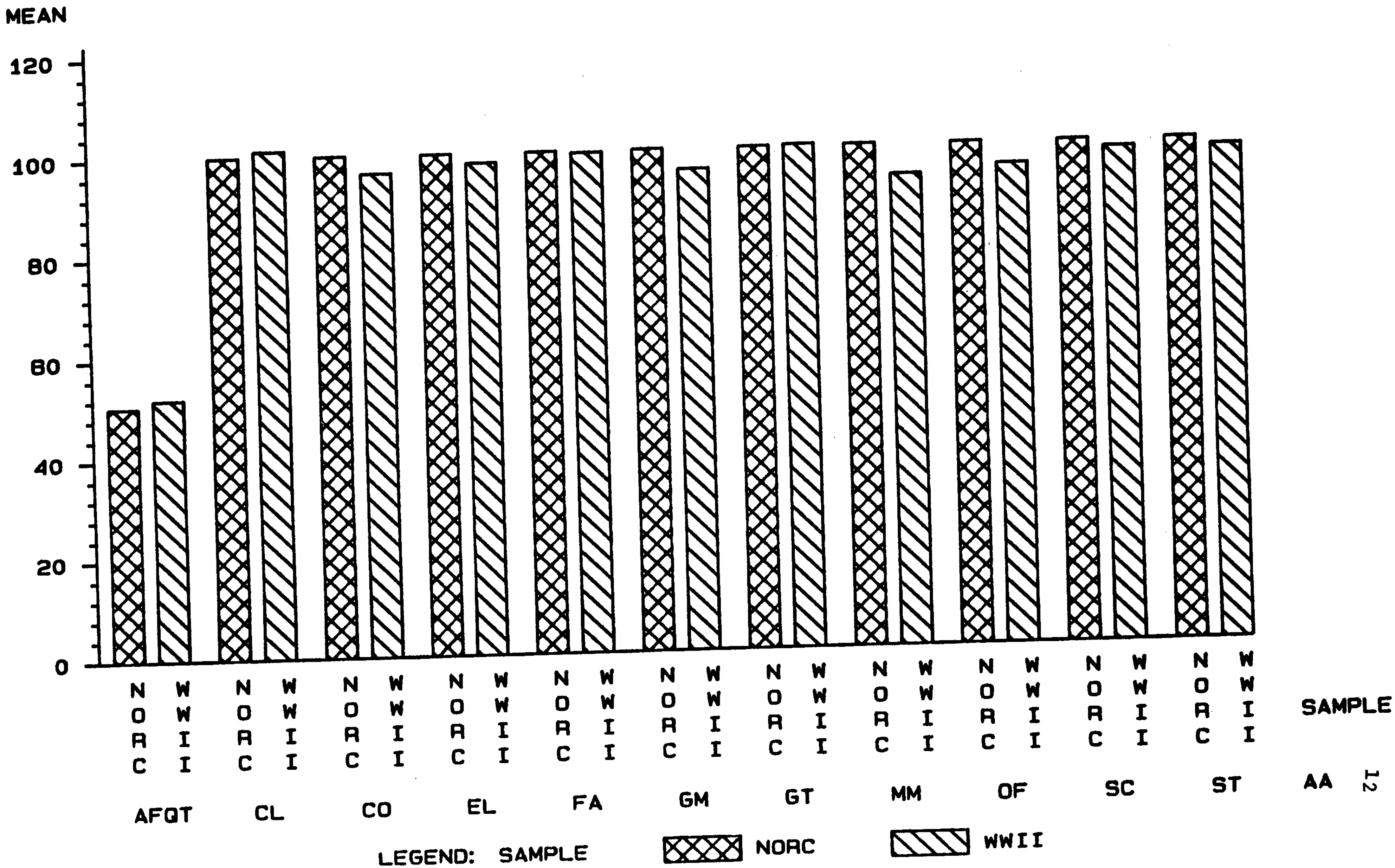
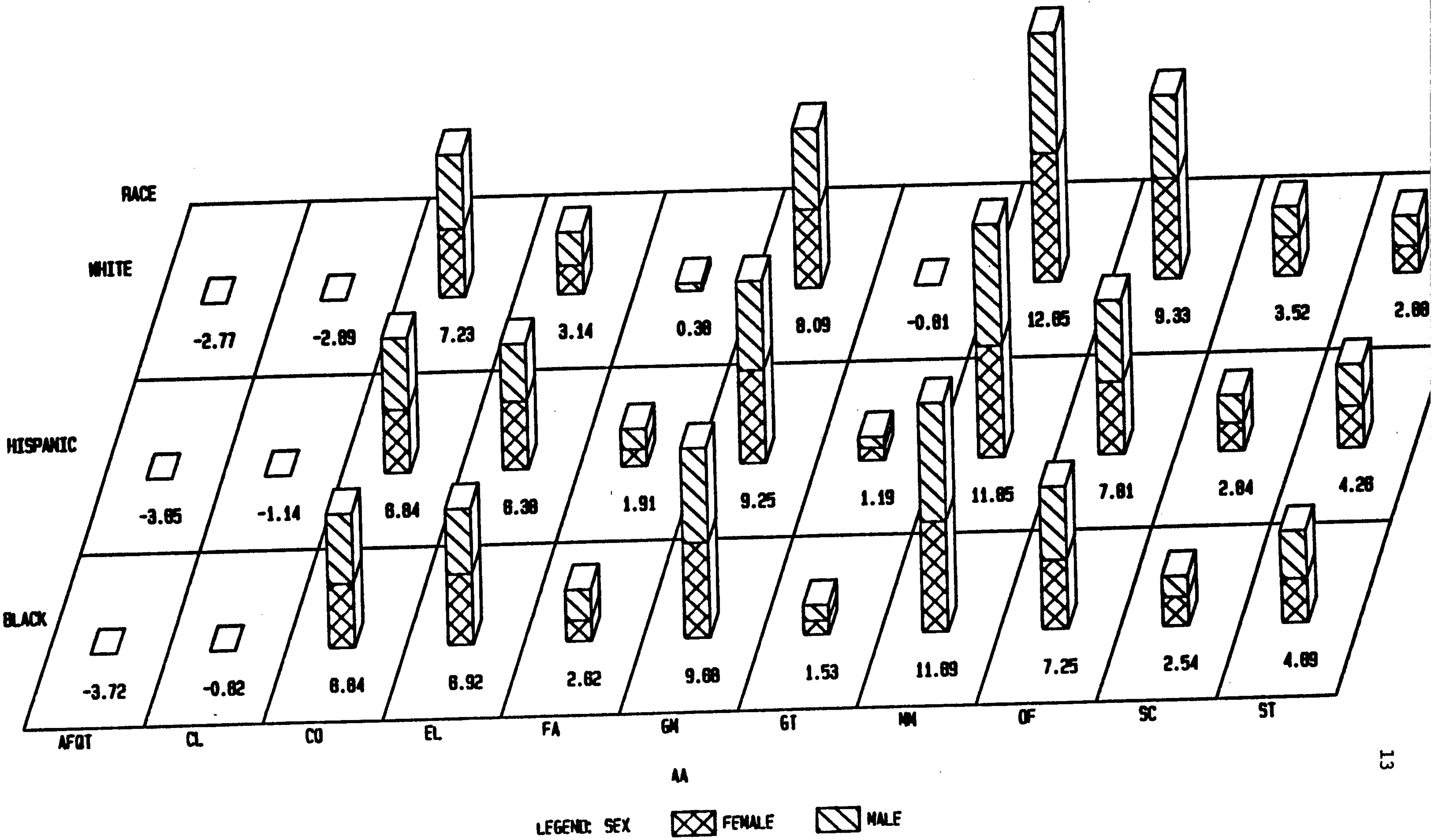


Figure 3
AVERAGE COMPOSITE DIFFERENCES
BY SUBGROUP

BLOCK CHART OF SUMS



Working Paper

RS-WP-84-13

STATUS REPORT ON IMPLEMENTATION OF THE 1980 YOUTH POPULATION NORMS FOR THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY, FORMS 11, 12, 13 AND 14 AND REVISION OF THE ARMY CLERICAL AND SURVEILLANCE/COMMUNICATIONS COMPOSITES

Karen J. Mitchell and Lawrence M. Hanser
SELECTION AND CLASSIFICATION TECHNICAL AREA

Frances C. Grafton
Personnel Policy Research Group

May 1984

Reviewed by
Newell K. Eaton
Selection and Classification
Technical Area

Approved by
Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

MEMORANDUM FOR:			DIRECTORATE	DAPE-MPA-CS
	ADCSPER		ACTION OFFICER	Mr. Ruberton
	DCSPER	X	TELEPHONE NO.	50836/30 April 1984

SUBJECT: Status Report on the Implementation of the 1980 Youth Population Norms for the Armed Services Vocational Aptitude Battery

☐ INFORMATION ☐ APPROVAL ☐ SIGNATURE ☐ INITIALS

SUSPENSE: SGS OSA OSD

COMMENT 1

FACTS:

a. Since 1944 the reference population for all military selection and classification instruments has been defined as all males serving in the military during World War II (WWII).

b. In February 1980 Congress questioned the appropriateness of interpreting the ability of today's youth in terms of the norms established on the World War II group. Documentation appears as Tab C.

c. In July 1980 OSD informed Congress that it was participating in the Profile of American Youth Study. The OSD research was designed to measure the vocational abilities of a nationally representative sample of 18-23 years olds and to develop a new reference population for the ASVAB. See documentation at Tabs C and D.

d. In October 1982 OSD decided to adopt the 1980 Youth Population as the new reference population for ASVAB beginning 1 October 1984.

FINDINGS:

a. By OSD definition, AFQT Categories I-IIIA should represent the top 50% of the military-eligible population.

b. Fifty-three percent of the 1980 Youth Population are classified in AFQT Categories I-IIIA when the World War II norms are applied (raw score of 76 or above). See Tab A and Army Research Institute Working Paper at Tab B.

c. The AFQT raw score which divides the 1980 Youth Population into the upper and lower 50 percent is 78.

COORDINATION	APPROVALS	INITIAL	DATE	ACTION TAKEN:	ADCSPER	DCSPER
PERI-RS	DIVISION CHIEF			CONCUR		
	DIRECTOR			APPROVED		
	EXECUTIVE			NOTED		
				SEE ME		
				BRIEF		

M\C ACTION: ☐ STAMP ☐ DISPATCH

SUBJECT: Status Report on the Implementation of the 1980 Youth Population Norms for the Armed Services Vocational Aptitude Battery

DISCUSSION:

a. In order to achieve an AFQT percentile score of 50 using the WWII norms an examinee is required to obtain a raw score of 76. Implementing the OSD decision to use the 1980 Youth reference population means that an examinee will have to have a raw score of 78 in order to achieve the same AFQT percentile score of 50. Similar adjustments have been made at each of the AFQT Category boundaries.

b. The transition from the WWII to the 1980 reference population also affects the Army Aptitude Area composite scores. In order to maintain current aptitude levels, DCSPER has redefined minimum aptitude area composite requirements for some MOS. These tables are at Tabs F and G.

c. In addition, the transition will affect examinees tested with ASVAB 5, 8, 9, and 10 who wish to enlist or reenlist on or after 1 October 1984.

IMPACT:

a. Fewer Army applicants will be classified in AFQT categories I-IIIA when the 1980 Youth Population norms are implemented. For example, 48.1% of FY83 Army applicants were classified in AFQT Categories I-IIIA using the World War II norms while 43.4% would have been in these categories using the 1980 Youth Population norms.

b. The aptitude area standards will have to be adjusted in Army MOS with CO, OF, GM, or MM aptitude area prerequisites. These tables are at Tab G.

c. Policies for all the Services pertaining to use of ASVAB 5, 8, 9, and 10 scores after 1 October 1984 are at Tab H.

RECOMMENDATION: Recommend the above changes in the Army Aptitude Area standards be approved. These changes are detailed at Tab G.

COORDINATION	APPROVALS	INITIAL	DATE	ACTION TAKEN:	ADCSPER	DCSPER
PERI-RS	DIVISION CHIEF			CONCUR		
	DIRECTOR			APPROVED		
	EXECUTIVE			NOTED		
				SEE ME		
				BRIEF		

M\C ACTION: ☐ STAMP ☐ DISPATCH

ODCSPER FORM 295-1 (SEPTEMBER 1970) CONTINUATION SHEET INTRAOFFICE CONTROL

MEMORANDUM FOR:			DIRECTORATE	DAPE-MPA-CS
	ADCSPER		ACTION OFFICER	Mr. Ruberton
	DCSPER	X	TELEPHONE NO.	50836/30 April 1984

SUBJECT: Implementation of Armed Services Vocational Aptitude Battery Forms 11, 12, 13, and 14 and Revision of the Army Clerical and Surveillance/Communications Composites

☐ INFORMATION ☐ APPROVAL ☐ SIGNATURE ☐ INITIALS

SUSPENSE: SGS OSA OSD

COMMENT 1

FACTS:

a. New forms of the Armed Services Vocational Aptitude Battery (ASVAB) will be implemented throughout MEPCOM on 1 October 1984. The new forms are ASVAB 11/12/13/14.

b. Two Army composites are recommended for revision on 1 October 1984. These are the Clerical (CL) and Surveillance/Communications (SC) composites.

c. The recommended composites are:

CL=VE+AR+MK

SC=VE+AR+AS+MC

These changes are supported by extensive research conducted for the DCSPER by the Army Research Institute. This research indicates that the recommended composites will yield better prediction of performance than the current composites. Additional information about the research appears at Tab E.

RECOMMENDATION: Recommend approval of the revised CL and SC composites described above.

COORDINATION	APPROVALS	INITIAL	DATE	ACTION TAKEN:	ADCSPER	DCSP
PERI-RS	DIVISION CHIEF			CONCUR		
	DIRECTOR			APPROVED		
	EXECUTIVE			NOTED		
				SEE ME		
				BRIEF		

M\C ACTION: ☐ STAMP ☐ DISPATCH

ODCSPER FORM 295
26 FEBRUARY 1969

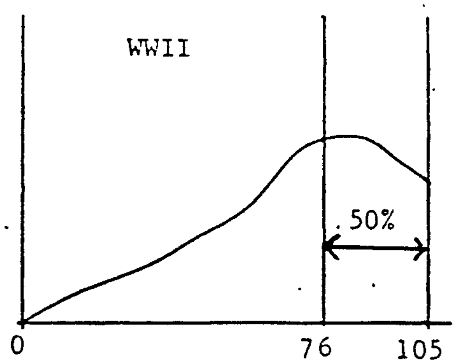
INTRAOFFICE CONTROL

PREVIOUS EDITION OF THIS FORM IS OBSOLETE.

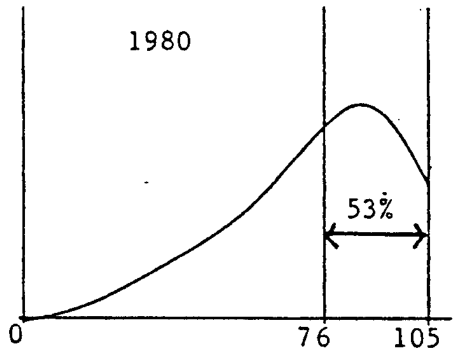
TAB A

CHANGING THE ASVAB REFERENCE POPULATION

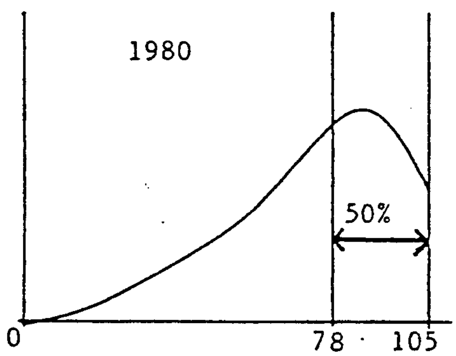
Distribution of AFQT Raw Scores



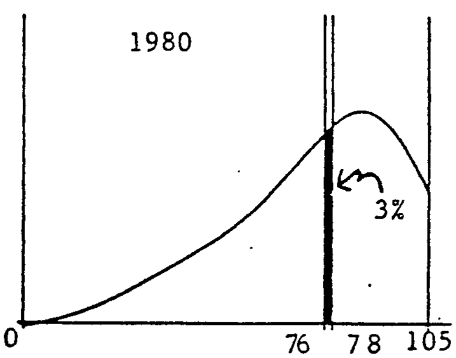
- TRADITIONAL DEFINITION OF AFQT CAT I-IIIA, AT OR ABOVE 50TH PERCENTILE (RAW SCORE 76)



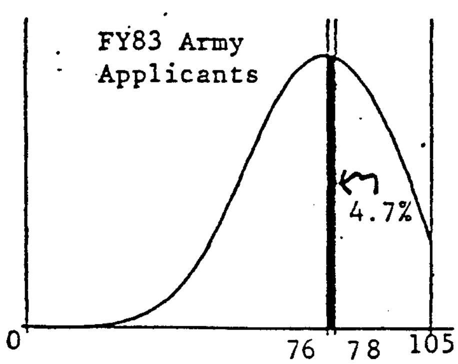
- OF THE 1980 REFERENCE POPULATION, 53% SCORE AT OR ABOVE AFQT RAW SCORE 76



- OF THE 1980 REFERENCE POPULATION, 50% SCORE AT OR ABOVE AFQT RAW SCORE 78



- OF THE 1980 REFERENCE POPULATION, 3% SCORE BETWEEN AFQT RAW SCORES 76 AND 78



- OF THE FY83 ARMY APPLICANTS, 4.7% SCORE BETWEEN AFQT RAW SCORES 76 AND 78

- 4.7% WERE AFQT CAT IIIA; WOULD BE AFQT CAT IIIB

TAB B

Working Paper

RS-WP-84-07

THE 1980 YOUTH POPULATION NORMS: ENLISTMENT AND OCCUPATIONAL CLASSIFICATION STANDARDS IN THE ARMY

Karen J. Mitchell and Lawrence M. Hanser
SELECTION AND CLASSIFICATION TECHNICAL AREA

May 1984

Reviewed by
Frances C. Grafton
Personnel Policy Research Group

Approved by
Newell K. Eaton
Chief
Selection and Classification Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

THE 1980 YOUTH POPULATION NORMS: ENLISTMENT AND OCCUPATIONAL CLASSIFICATION STANDARDS IN THE ARMY

Karen J. Mitchell
Lawrence M. Hanser

INTRODUCTION

Since 1944 the cognitive ability of military applicants has been referenced statistically to the population of adult males tested during World War II. The Department of Defense and Congress have recently questioned the appropriateness of using the norms established on the World War II reference population as a metric for interpreting the ability of today's applicants. In 1979 a decision was made to examine the distribution of cognitive ability of today's 18-23 year olds. An understanding of the distribution of cognitive ability of the current military-eligible population would allow policy makers to establish entrance standards and accession quotas that are more compatible with present manpower resources.

In 1980 the Department of Defense, in cooperation with the Department of Labor, sponsored the Profile of American Youth Study. This study was designed to assess the vocational aptitudes of a nationally representative sample of 18-23 year olds and to develop new national norms for the Department of Defense enlistment test, the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB is composed of eight power and two speed tests. Four ASVAB subtests are combined to form the Armed Forces Qualification Test (AFQT), a general measure of trainability and the primary criterion for enlistment. Other subtest combinations are used in occupational classification. A complete discussion of the content and structure of ASVAB may be found in Kass, Mitchell, Grafton & Wing (1982).

The National Opinion Research Center (NORC) of the University of Chicago contracted with the Department of Defense to administer ASVAB from July through October 1980 to a nationally representative sample of approximately 12,000 men and women. The sample was constructed to represent age/sex/ethnicity variables and major census regions. The data were used to compare the 1980 youth population to the World War II reference population and to FY81 military accessions. Subgroup comparisons were conducted on the basis of age, sex, race/ethnicity, level of education, socioeconomic status, and geographic region. This work is described in "Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery," March 1982. The data were also used to establish and verify new score scales for the ASVAB subtests

and aptitude area composites (Maier & Sims, 1982; Ree, Valentine & Earles, 1983).

Early in 1983 Sims and Maier reported apparent score pattern discrepancies on the speeded subtests between the male NORC youth sample and a series of male military applicants. The discrepancy resulted in lower scores for the 1980 sample males on the two speeded subtests when compared to a sample of male military examinees. An analysis of possible sources of the discrepancy revealed that there were format differences between answer sheets used in the NORC field test and operational answer sheets used for military entrance testing.

In August of 1983, Earles, Guiliano, Ree & Valentine outlined a plan for estimating the differences in performance on the speeded subtests attributable to the non-standard NORC answer sheet. A study was conducted in the fall of 1983 at the Military Enlistment Processing Stations to develop corrections for the NORC speeded subtests and to adjust the percentile conversion tables. The corrected subtest and composite norms became available to the Service Labs in January of 1984. They will be implemented in October 1984 with the introduction of ASVAB forms 11/12/13.

PROBLEM

Since 1944 various forms of AFQT and ASVAB have been calibrated to the World War II mobilization population. The use of a constant reference population allows the Services to (1) track the quality of applicants and recruits over time, and (2) maintain enlistment and occupational classification standards across test forms and over time. With the introduction of a new normative base for ASVAB, research is needed to document the comparability of applicant and accession quality based on the World War II and NORC reference groups. Data are also needed on the appropriateness of maintaining current enlistment and occupational classification standards after October 1984.

METHOD

Scores of the NORC sample on the power subtests were weighted and converted to subtest standard scores using both the World War II and NORC norms. The NORC sample raw scores on the speeded subtests were converted to equated raw scores based on the January 1984 tables and then converted to subtest standard scores using both sets of norms. This procedure yielded two sets of subtest standard scores for the NORC sample: (1) standard scores referenced to the WWII population and (2) standard scores referenced to the 1980 Youth Population. Mean subtest standard scores were computed for these two sets of scores.

The subtest standard scores were then combined to form the ten Army aptitude area composites. These standard score sums were converted to Army

standard scores using the World War II and NORC norms. Mean composite scores were computed for each. Subtest raw scores for Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), and Numerical Operations (NO) subtests were also summed and the sum converted to percentiles for the AFQT index. Mean AFQT scores were computed using both norms.

The composite scores were computed and examined for each of six sex/ethnicity subgroups: (1) white males, (2) black males, (3) Hispanic males, (4) white females, (5) black females, and (6) Hispanic females. Finally, the scores at each of the current occupational classification cutpoints for the ten aptitude areas were examined in order to compare the differences between scoring tables based on the two different norm groups. To the extent that the ability scores assigned by the new norms differed from the ability scores assigned on the basis of the historical norm group, questions of adjusting the current enlistment and occupational classification standards would arise.

RESULTS AND DISCUSSION

Mean subtest standard scores using the NORC and World War II norms appear at Table 1. This information is reported graphically at Figure 1. The mean NORC-assigned verbal, math, and speeded subtest scores were lower than the World War II-converted means. This means, for example, that an examinee receiving a Numerical Operations (NO) score of 50 using the new norms would have received a 51 in terms of the World War II subtest norms. An examinee receiving a 50 on Mathematics Knowledge (MK) using the new norms would have received a 52 using the historical conversions.

The converse is true for the technical subtests, Auto-Shop Information (AS), Mechanical Comprehension (MC), and Electronics Information (EI). The assigned NORC score was from two to four points higher than the World War II-assigned score. An applicant receiving a score of 50 on AS would have received a score of 46 using the World War II norms. The mean General Science (GS) scores using the two conversions were roughly equivalent. Mean aptitude area composite scores appear at Table 2. The data are presented graphically at Figure 2. The mean NORC-assigned composites for Clerical (CL), General Technical (GT), and AFQT were lower than the traditional norm scores. Thus, an applicant receiving a score at the 51st percentile on AFQT using the 1980 norms would be at the 52nd percentile on the WWII norms. For the remainder of the composites, the NORC-assigned composite score exceeded the World War II-converted score. The mean aptitude area differences ranged from one half of an Army Standard Score point on Field Artillary (FA) to six standard score points on Mechanical Maintenance (MM).

Table 3 presents the mean aptitude area score differences for the sex/race subgroups. The entries are the differences between the NORC- and World War II-assigned aptitude area scores. In all cases the historical

score is the subtrahend. The differences are presented pictorially at Figure 3. Examination of differences across the rows revealed no apparent subgroup patterns.

Table 4 contains the NORC aptitude area cutoff scores that would have to be implemented to maintain ability levels equivalent to the enlistment and occupational classification standards that are currently in place. Table 5 provides the same data with the NORC aptitude area cutoff scores rounded to five-point score intervals.

REFERENCES

Earles, James A., Giuliano, Toni 1LT, Ree, Malcolm J. & Valentine, Lonnie D. (1983). The 1980 Youth Population: An Investigation of Speeded Subtests. Brooks, AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Kass, Richard A., Mitchell, Karen J., Grafton, Frances C. & Wing, Hilda (1982). Factor Structure of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10: 1981 Army Applicant Sample. Alexandria, Virginia: Army Research Institute, Technical Report 581.

Maier, M. H. & Sims, W. H. (1982). Constructing an ASVAB Score Scale in the 1980 Reference Population. Alexandria, Virginia: Center for Naval Analysis, Memorandum 82-3188.

Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery (1982). Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Ree, M. J., Valentine, L. D., Jr., & Earles, J. A. (1983). The 1980 Youth Population: A Verification Report. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Sims, W. H. & Maier, M. H. (1983). The Appropriateness for Military Applications of the ASVAB Subtests and Score Scale in the new 1980 Reference Population. Alexandria, Virginia: Center for Naval Analyses, Memorandum 83-3102.

Table 1

MEAN SUBTEST STANDARD SCORES

NORC and WWII Norms

Subtest	NORC		WWII	
	$\overline{X}$	SD	$\overline{X}$	SD
GS	49.90	10.01	49.63	9.70
AR	49.96	10.01	50.33	10.25
WK	49.98	9.96	50.81	10.05
PC	50.04	10.04	51.47	9.66
NO	50.04	9.98	51.14	10.33
CS	49.98	10.01	52.77	10.39
AS	49.97	10.00	46.26	99.23
MK	50.01	9.99	51.84	10.77
MC	50.00	10.04	47.55	9.55
EI	49.97	10.00	47.98	9.86

Table 2

MEAN APTITUDE AREA COMPOSITE SCORES

NORC AND WWII NORMS

AA	NORC		WWII		SUBTESTS
	$\overline{X}$	SD	$\overline{X}$	SD	
CO	100.00	20.01	96.4	19.94	AR, AS, MC, CS
FA	100.01	20.00	99.58	20.53	AR, MK, MC, CS
EL	99.99	20.00	98.08	21.00	AR, EI, MK, GS
OF	100.0	20.00	95.55	19.89	NO, VE, MC, AS
SC	99.99	20.00	98.37	20.16	NO, CS, VE, AS
MM	100.00	19.99	93.78	20.06	NO, EI, MC, AS
GM	100.0	20.00	95.81	20.49	MK, EI, GS, AS
CL	100.0	20.01	101.24	21.39	NO, CS, VE
ST	99.99	20.00	98.39	21.71	VE, MK, MC, GS
GT	100.0	20.01	100.19	21.41	VE, AR
AFQT	50.65	29.14	52.13	28.29	VE, AR, NO

NOTE: The CL and SC aptitude area composites will undergo revision when ASVAB 11/12/13 is implemented. Results reported here pertain to the 8/9/10 CL and SC composites

Table 3

MEAN APTITUDE AREA DIFFERENCES

BY RACE/SEX SUBGROUPS

AA	WHITE MALES	BLACK MALES	HISPANIC MALES	WHITE FEMALES	BLACK FEMALES	HISPANIC FEMALES
CO	3.82[a]	3.61	3.67	3.45	3.29	3.25
FA	.53	1.59	1.12	-.08	1.11	.91
EL	1.69	3.34	2.95	1.42	3.56	3.41
OF	4.19	3.71	4.15	5.08	3.52	3.65
SC[b]	1.53	1.03	1.38	1.88	1.45	1.37
MM	6.11	6.04	6.16	6.51	5.65	5.70
GM	4.12	4.84	4.49	3.96	4.87	4.77
CL[b]	-1.07	-.48	-.48	-1.81	-.28	-.65
ST	1.53	2.42	2.18	1.34	2.24	2.13
GT	-.57	.80	.58	-.23	.68	.59
AFQT	-1.28	-1.86	-1.79	-1.49	-1.86	-1.86

[a] NORC - WWII

[b] NOTE: The CL and SC aptitude area composites will undergo revision when ASVAB 11/12/13 is implemented. Results reported here pertain to the 8/9/10 CL and SC composites:

Table 4

Equivalence of NORC and WWII Aptitude Area Score Cutoffs

WWII Cut Off		80	85	90	95	100	105	110	115	120
	AA / CO	84	88	93	98	103	110	114	119	123
	FA	81	85	89	94	99	104	109	115	121
	EL	83	87	90	96	100	105	110	116	122
	OF	85	90	95	101	106	111	115	119	122
NORC	SC[a]	81	85	89	94	99	103	108	114	119
CUTOFF	MM	87	92	96	102	108	113	116	121	125
	GM	84	89	94	99	104	110	114	118	124
	CL[a]	81	84	88	92	99	103	108	114	120
	ST	82	87	90	96	101	107	110	116	122
	GT	81	86	90	96	101	105	110	115	120

[a]NOTE: The CL and SC aptitude area composites will undergo revision when ASVAB 11/12/13 is implemented. Results reported here pertain to the 11/12/13 CL and SC composites. The new composites are: CL=VE+AR+MK, and SC=VE+AR+AS+MC

Table 5

Equivalence of Rounded NORC and WWII Aptitude Area Score Cutoffs

WWII Cut Off		80	85	90	95	100	105	110	115	120
	AA/CO	85	90	95	100	105	110	115	120	125
	FA	80	85	90	95	100	105	110	115	120
	EL	85	85	90	95	100	105	110	115	120
	OF	85	90	95	100	105	110	115	120	120
NORC	SC[a]	80	85	90	95	100	105	110	115	120
CUTOFF	MM	85	90	95	100	110	115	115	120	125
	GM	85	90	95	100	105	110	115	120	125
	CL[a]	80	85	90	95	100	105	110	115	120
	ST	80	85	90	95	100	105	110	115	120
	GT	80	85	90	95	100	105	110	115	120

[a]NOTE: The CL and SC aptitude area composites will undergo revision when ASVAB 11/12/13 is implemented. Results reported here pertain to the 11/12/13 CL and SC composites. The new composites are: CL=VE+AR+MK, and SC=VE+AR+AS+MC

Figure 1
MEAN SUBTEST SCORES
NORC AND WWII NORMS

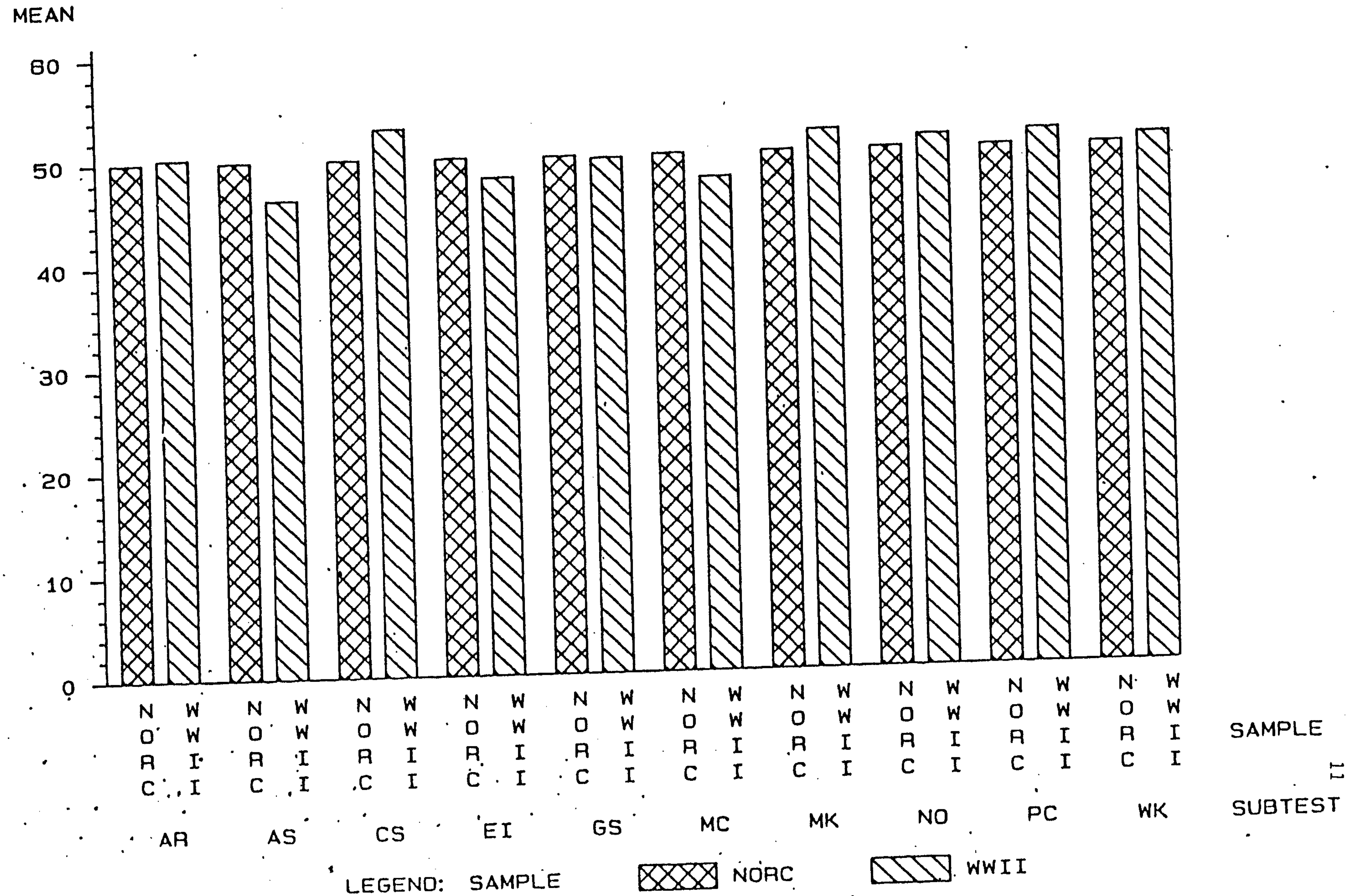


Figure 2

MEAN APTITUDE AREA SCORES
NORC AND WWII NORMS

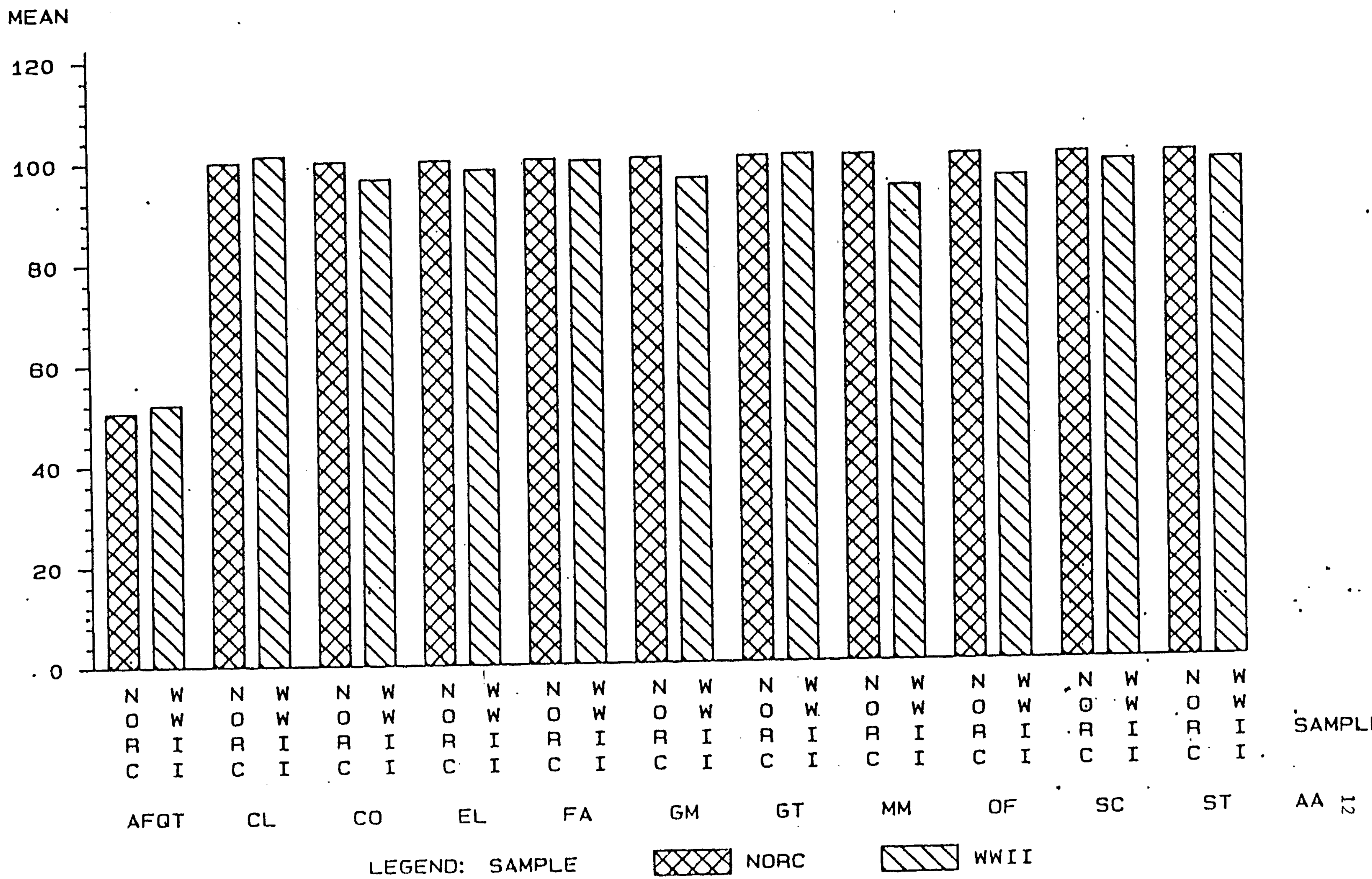
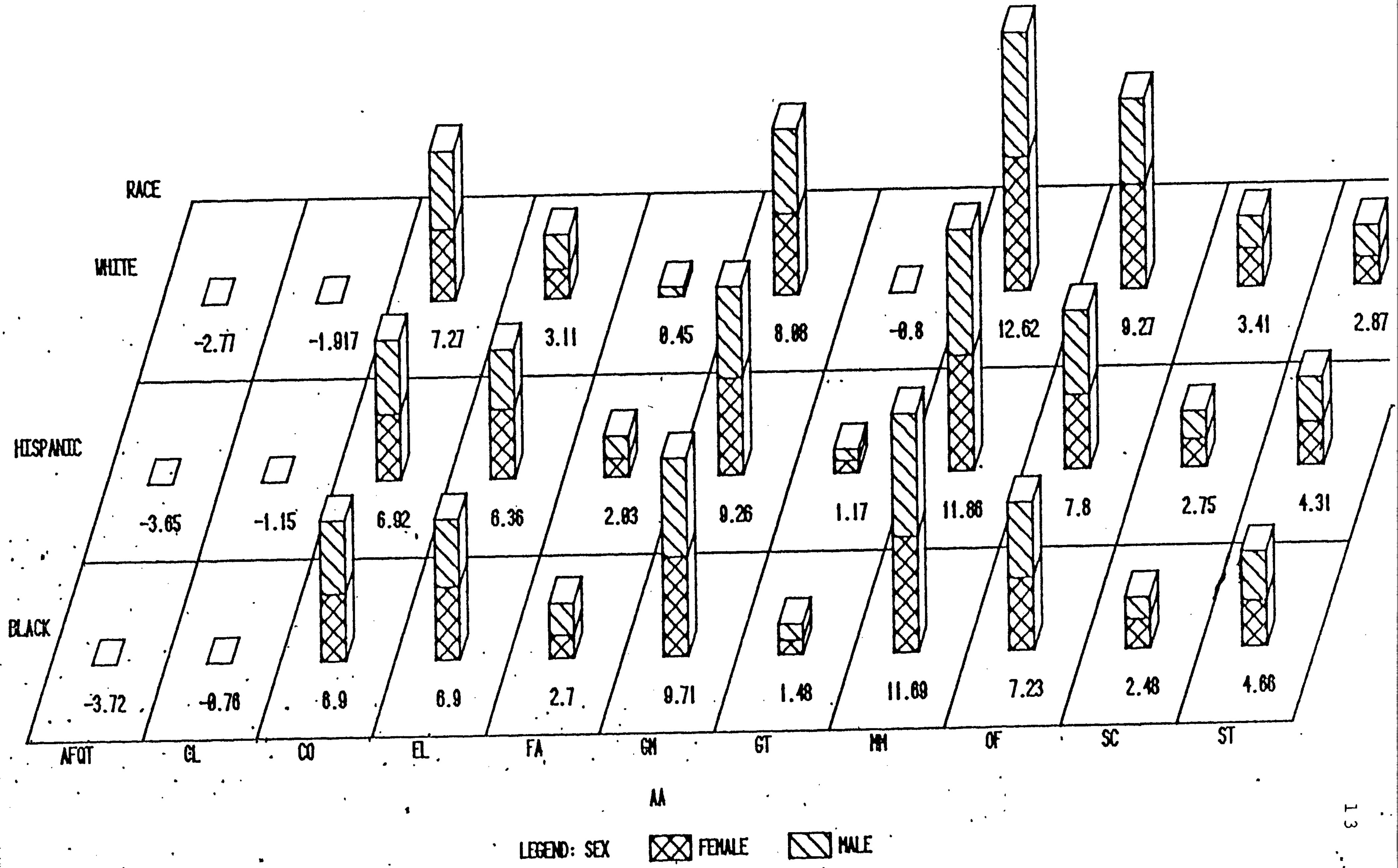


Figure 3
AVERAGE COMPOSITE DIFFERENCES
BY SUBGROUP

BLOCK CHART OF SUMS



TAB C

EXTRACT FROM OASD (M,RA&L)
MEMORANDUM FOR RECORD, DATED 15 DEC 82
PERTAINING TO THE ASVAB STEERING COMMITTEE
MEETING OF 26 OCT 82

Development of 1980 Reference Population. Dr. W. S. Sellman introduced this topic by providing historical background. When the Department of Defense (DoD) informed the Congress in February 1980 about the ASVAB calibration problem, the Armed Services Committees were incredulous to learn that enlistment test scores were interpreted against a reference population of all men (officers and enlisted personnel) under arms as of December 31, 1944. The assumption in Congress was that the test scores of new recruits were compared with the scores of contemporary youth — that is, the most current nationwide pool of potentially eligible enlistees.

Some Congressmen, in reaction, thought that the delay between the time that DoD first suspected the miscalibration in 1976 and the official testimony in 1980 describing it was part of a purposeful deception. They maintained that DoD knew all along about the updated test results, and it deliberately concealed the statistical error in an effort to make the All-Volunteer Force (AVF) look better than it was in reality.

Another group of legislators — supporters of the AVF — thought the timing of DoD's revelation coincided too perfectly with the Carter Administration's push for return to draft registration. The admission of error by DoD testing experts, and the consequent enlistment of otherwise unqualified recruits, was seen as an attempt to buttress the support for the registration of men and women.

The integrity of the DoD testing program, among both the supporters and detractors of the AVF, was badly undermined. In an attempt to restore some of its credibility and to solve the problem of obtaining a current reference population, DoD administered the ASVAB to a nationally representative sample of young people as part of the "Profile of American Youth" study. This, it was hoped, would establish once and for all whether the kinds of people entering the military were reasonably representative of the general population, or the poorly educated, low-ability individuals so frequently depicted in the media. In the 1980 and 1981 Defense Manpower Overview testimony and in a July 1980 report to the House Committee on Armed Services entitled, "Aptitude Testing of Recruits", DoD advised the Congress that it was conducting the "Profile" study and developing a new reference population for interpreting enlistment test scores. (CONTINUED ON NEXT PAGE)

Note:
Told Congress "developing a new ref pop for int and test scores" — means it is being done! CNW

Encl. 1

EXTRACT FROM OASD (M,RA&L)
MEMORANDUM FOR RECORD, DATED 15 DEC 82
PERTAINING TO THE ASVAB STEERING COMMITTEE
MEETING OF 26 OCT 82

At its 26-28 January 1982 meeting, the ASVAB Working Group discussed technical aspects of using the "Profile" data to define a contemporary reference population for interpreting ASVAB scores. Among others, these included issues such as the location of AFQT category boundaries, and the impact of a new reference population on Service qualification rates for enlistment and job assignment. Obviously, development of a new reference population has serious implications for manpower planning and should be accomplished only after careful analysis of both policy and psychometric considerations.

In a 27 April 1982 memorandum, Dr. Sicilia asked the Marine Corps to undertake the construction of new norms for ASVAB using the 1980 reference population data. In particular, Drs. M. H. Maier and W. H. Sims, Center for Naval Analyses, were asked to develop score conversions for AFQT and aptitude composites with special emphasis on how a new reference population would affect current enlistment standards and the proportions of new recruits who would be placed in the various AFQT categories.

Following Dr. Sellman's introduction, Dr. M. H. Maier presented a briefing on behalf of the Working Group which recommended changing the ASVAB score scale from the 1944 reference population to the 1980 youth population. (A copy of Dr. Maier's briefing is at attachment 3.) In particular, Dr. Maier recommended that the Steering Committee (1) adopt the 1980 reference population, (2) define the 1980 reference population to consist of 18 to 23-year old males and females, and (3) define AFQT category bounderies to maintain constant percentile scores (i.e., the same boundaries as currently exist). Reasons for recommending the change included: (1) the 1944 reference population is almost 40 years old and cannot be statistically reproduced, (2) because of changes in AFQT content, length, and administrative instructions, DoD's capability to track current test scores back to that reference population is problematic, and (3) the continued use of the 1944 reference population precludes the comparison of the aptitudes of new recruits with those of the manpower pool from which they are drawn. The recommendations presented by Dr. Maier were also strongly endorsed by the Defense Advisory Committee on Military Personnel Testing.

Finally, Dr. Maier showed the scores from the 1980 scale that would be required to maintain the current enlistment standards: AFQT scores would require either no change or a one-point change. Aptitude composite scores for the Army and Air Force would need to be raised somewhat.

TAB D

EXTRACT FROM JULY 1980 OASD (M,RA&L)
REPORT TO THE HOUSE COMMITTEE
ON ARMED SERVICES

CORRECTING THE CALIBRATION PROBLEM

The testing experts employed by the Department of Defense and the outside consultants hired by the Department now agree on the adjustments that should be made to the scoring system which converts raw scores to percentile scores. The Department of Defense began working on new versions of the ASVAB about two years ago as part of its continuous effort to improve the test and provide fresh forms. The new forms will be correctly calibrated when they are introduced. In addition, they are designed to be better predictors of performance in military service. As an added benefit, the availability of the new forms will reduce the opportunity for test compromise.

In addition to improving the Department's capability to make historic comparisons, work is also underway to administer the new ASVAB to a representative sample of the nation's young men and women. This sample was developed by the Department of Labor. The tests will be administered by the National Opinion Research Center of the University of Chicago. When the results are available in early 1981, the Department will have, for the first time, the means to compare new enlistments with the current youth population as well as the 1944 mobilization group. This information is essential for managing recruiting and making realistic judgments on recruiting results. The Department recruits from today's population, not the 1944 one. The information is also needed for mobilization planning. The Department needs to compare the characteristics of the current youth population to its requirements for wartime manpower. If a national emergency required the resumption of the draft, the Department should be in a position to recommend a cut-off score on the entrance test that is fair and meets the manpower needs of the Military Services. Decisions on who should be drafted, no less than on who should be permitted to volunteer, require a knowledge of the characteristics of the current youth population.

Positive Statement

WW

TAB E

ARI COMMENTS ON EVALUATION OF ARMY
AREA COMPOSITES

Two Army composites are recommended for revision on
~~The composition of two of the current Army composites require~~
~~change with the~~ 1 October 1984. ~~implementation of ASVAB 11/12/13.~~
These are ~~T~~he Army Clerical (CL) and Surveillance/Communications (SC)
composites. ~~will be revised; the new structures are:~~ The recommended composites are:

CL = VE+AR+MK, and

SC = VE+AR+AS+MC.

The change is supported by extensive research conducted for ODCSPER by the Army Research Institute (ARI). Part of the research addresses the predictive validity of the ASVAB 8/9/10 Army Aptitude Area Composites. Analyses indicate that the revised CL and SC composites listed above yield better prediction of performance in training and Army jobs than do the 8/9/10 compositions. While the research found no evidence to support the use of as many as the ten composites currently used by the Army, with the proposed changes in CL and SC, the new system will perform as well as a revised system with fewer composites. Further, ARI is continuing in the long-term Soldier Selection and Classification Research (Project A) to evaluate new psychomotor, perceptual, interest, cognitive, and temperament measures (which may later be added to the entrance battery). Given the difficulties in changing the current ten-composite system and likely substantial improvements from the ongoing ARI research project, ARI recommends keeping the current composite system with revisions in CL and SC as indicated above. A report documenting the analyses is scheduled for release by ARI on 15 May 1984.

Pls check this. I made a few changes or circled. V.

TAB F

EQUIVALENCE OF ROUNDED NORC AND WWII
APTITUDE AREA SCORE CUTOFFS

WWII CUT OFF		80	85	90	95	100	105	110	115	120
	AA									
	CO	85	90	95	100	105	110	115	120	125
	FA	80	85	90	95	100	105	110	115	120
	EL	85	85	90	95	100	105	110	115	120
	OF	85	90	95	100	105	110	115	120	120
NORC	SC	80	85	90	95	100	105	110	115	120
CUT OFF	MM	85	90	95	100	110	115	115	120	125
	GM	85	90	95	100	105	110	115	120	125
	CL	80	85	90	95	100	105	110	115	120
	ST	80	85	90	95	100	105	110	115	120
	GT	80	85	90	95	100	105	110	115	120

NOTE: The CL and SC aptitude area composites will undergo revision when ASVAB 11/12/13 is implemented. Results reported here pertain to the revised 11/12/13 CL and SC composites. The new composites are: CL=VE+AR+MK, and SC=VE+AR+AS+MC.

TAB D

TAB G

CHANGES IN APTITUDE AREA SCORES

CO 85 to 90	90 to 95	95 to 100	100 to 105	105 to 110	110 to 115	115 to 120	120 to 12
COMBAT							
11B1		12E1					
11C1							
11H1							
11M1							
12B1							
12C1							
12F1							
19D1							
19E1							
19K1							
19A1							

OF 85 to 90	90 to 95	95 to 100	100 to 105	105 to 110	110 to 115	115 to 120	120 to 12
OPR & FOOD							
16F1	16T1	15D1	13M1				
16S1		15E1					
64C1		16B1					
94B1		16C1					
		16D1					
		16E1					
		16H1					
		16J1					
		16L1					
		16P1					
		16R1					
		94F1					

GM 85 to 90	90 to 95	95 to 100	100 to 105	105 to 110	110 to 115	115 to 120	120 to 12
GEN MAINT							
41J1	41C1	41B1	00B1				
43E1	45T1	42C1	55D1				
43M1	53B1	42E1					
44B1	55B1	44E1					
45B1	62G1	45D1					
51B1	68M1	45K1					
51C1		45L1					
51K1		51G1					
51M1		52C1					
51N1		52D1					
57E1		52F1					
57F1		54C1					
57H1		55G1					
61F1		68J1					
62E1							
62F1							
62H1							
62J1							

*SC 85 to 85	90 to 90	95 to 95	100 to 100	105 to 105	110 to 110	115 to 115	120 to 1
SURV/COMMO							
	05B1	17C1	05C1				
	31K1	17K1	13R1				
	72E1	17L1	17B1				
	72G1	17M1					
		96H1					

*No changes in score levels between WWII and 1980 youth norms.

*FA 85 to 85	90 to 90	95 to 95	100 to 100	105 to 105	110 to 110	115 to 115	120 to 12
FIELD ARTILLERY							
13B1			13F1				
			15J1				
*ST							
SKILLED TECH							
03C1		05D1	55R1	71Q1	00J1	33S1	
81C1		05G1	74D1	71R1	92D1		
83E1		05H1	74F1	73D1			
83F1		05K1	81Q1	91G1			
84C1		13C1	91P1	96B1			
95C1		13E1	91R1	97B1			
		42D1	93H1	98C1			
		54E1	93J1	98J1			
		71P1	95B1				
		81B1					
		81E1					
		82B1					
		82C1					
		82D1					
		84B1					
		84F1					
		91A1					
		91B1					
		91C1					
		91D1					
		91E1					
		91F1					
		91H1					
		91J1					
		91L1					
		91N1					
		91Q1					
		91S1					
		91T1					
		91U1					
		91V1					
		91Y1					
		92B1					
		92C1					
		93E1					
		93F1					
		96C1					
		96D1					
		98G1					

*No changes in score levels between WWII and 1980 youth norms.

*CL 85 to 85	90 to 90	95 to 95	100 to 100	105 to 105	110 to 110	115 to 115	120 to 12
CLERICAL							
	75P1	71C1					
	76V1	71G1	71N1	75F1	71D1		
	76W1	71L1					
		71M1					
		73C1					
		74B1					
		75B1					
		75C1					
		75D1					
		75E1					
		76C1					
		76J1					
		76Y1					
*EL 85 to 85	90 to 90	95 to 95	100 to 100	105 to 105	110 to 110	115 to 115	120 to 12
ELECTRONICS							
25J1	36C1	21G1	23N1	22L1	21L1	26C1	26Y1
41G1	36D1	22N1	26L1	24E1	24C1	31S1	35H1
	36E1	23U1	27F1	24W1	24G1		
	36K1	24L1	27P1	26D1	24H1		
		25L1	27Q1	31T1	24J1		
		26B1	31VS	32D1	24K1		
		26H1	32G1	46N1	24M1		
		26M1	32H1		24N1		
		26N1	35L1		24P1		
		26Q1	35M1		24Q1		
		26R1	35R1		24U1		
		26T1	36H1		26E1		
		27B1	36M1		26F1		
		27E1			26K1		
		27G1			26V1		
		27H1			27N1		
		27L1			31E1		
		27M1			31J1		
		31M1			32F1		
		31N1			34C2		
		31V1			34E2		
		34B1			34F2		
		35E1			34H2		
		35F1			34L1		
		35K1			34Y1		
		41E1			35C1		
		45G1			35G1		
		51R1			36L1		
		52G1					
		93F1					

*No changes in score levels between WWII and 1980 youth norms.

MM	85 to 90	90 to 95	95 to 100	100 to 110	105 to 115	110 to 115	115 to 120	120 to 12
MECH MAINT								
	62B1		45E1	24T1				
	63B1		42N1	61C1				
	63H1		61B1	63D1				
	63J1		63E1	63G1				
	63W1		63N1	63S1				
				63T1				
				63Y1				
				65B1				
				65D1				
				65E1				
				65F1				
				65H1				
				65J1				
				67G1				
				67H1				
				67N1				
				67T1				
				67U1				
				67V1				
				67X1				
				67Y1				
				68B1				
				68D1				
				65F1				
				68G1				
				68H1				
	(5)		(5)	(26)				

TAB H

5,8,9,10 ASVAB

Introduction of ASVAB 1/12/13

USE OF ~~TEST~~ SCORES AFTER 1 OCT 84
~~(When new forms of ASVAB are introduced)~~

1. After 30 September 1984, all applicants in the DEP will continue to be qualified for enlistment based upon test scores prior to 1 October 1984. Their scores will be converted to the 1980 youth reference population by MEPCOM for accession reporting purposes,

2. Applicants tested prior to 1 October 1984, who did not enlist in the DEP and desire to enlist after that date will have their test scores converted to the new 1980 youth reference norms. If they remain qualified, regular enlistment procedures apply. If they fail to qualify, an immediate retest is authorized which become the scores of record.

3. All ASVAB test scores for individuals tested with any form of ASVAB (5, 8, 9, or 10) are valid for two years from original test date. They will be converted to the 1980 norms. Provisions of paragraph 2 above will apply for applicants who fail to qualify.

4. In July 1984, a new form of ASVAB (#14) will be introduced solely for testing high school students. This will replace ASVAB 5 which has been used since 1976.

5. The scores from ASVAB 14 will be based upon the 1980 youth population norms and will be available for enlistment purposes on 1 October 1984. They cannot be used before 1 October 1984 because the new 1980 norms for the reference population cannot be used before that time.

6. Students tested during July through September 1984, who desire to enlist prior to 1 October 1984, will have to be retested with the production ASVAB test.

7. Prior service applicants enlisting in the same MOS in which they previously served will be acceptable, using previous scores which are valid for 5 years in accordance with AR 601-210. Their scores will be converted to the 1980 norms for reporting purposes.

- Those applying for a different MOS must have their previous test scores converted to the 1980 reference base, or retest if they are no longer qualified.

- Those applying for a Service other than the one in which they previously served, must be retested.

Working Paper

RS-WP-84-23

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB-PERFORMANCE

Karen J. Mitchell
SELECTION AND CLASSIFICATION TECHNICAL AREA

Jim H. Harris
HumRRo

October 1984

Reviewed by
for Lawrence Hanser
Selection and Classification
Technical Area

Approved by
for Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

PP 6742

ARMY RESEARCH TO LINK STANDARDS FOR ENLISTMENT TO ON-THE-JOB-PERFORMANCE

ARMY RESEARCH OBJECTIVES

OVERALL ARMY GOALS

The Army Research Institute is currently engaged in a large-scale, multi-year project to improve the Army selection and classification system and, thereby, increase the overall effectiveness of the force. The goal of the Army's program for increasing the efficiency of enlisted personnel selection and utilization is to enable the Army to meet its peacetime and mobilization missions through improved matching of individuals to military occupational specialties (MOS). The research is aimed at developing comprehensive selection and classification procedures to validly predict performance in Army training and occupational specialties. Specifically, this project will:

1. Validate existing selection measures against both existing and project-developed criteria. The criteria will include both Army-wide performance measures based on newly developed rating scales and direct measures of MOS-specific task performance.

2. Develop and validate new and/or improved selection and classification measures.

3. Validate intermediate criteria, such as performance in training, as predictors of subsequent criteria, eg. job performance ratings, so that informed reassignment and promotion decisions can be made throughout an individual's tour.

4. Determine the relative utility to the Army of different performance levels across MOS.

5. Estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making operational selection and classification decisions.

A complete description of the Army's research program and second year accomplishments have been published separately by the U.S. Army Research Institute for the Behavioral and Social Sciences in the project annual report for "Improving the Selection, Classification and Utilization of Army Enlisted Personnel" (1983).

A detailed description of the project also appears in the Second Annual Report to Congress on Joint-Service Efforts to Link Standards for Enlistment to On-the-Job Performance (1983).

Joint Project Goals

The Army goals and the joint project goals are the same.

SPECIALTIES SELECTED FOR JOINT STUDY

The Army's research focuses on 19 MOS. The MOS were selected to be representative of the Army and include all operational ASVAB aptitude area composites. Blacks, whites, Hispanics, males, and females are present in these MOS in the same proportions as in total accessions. These MOS represent 44% of annual Army enlistments.

A number of performance measures, including measures of training success, service-wide performance, and MOS-specific hands-on performance, will be developed for these MOS. For reasons of cost efficiency, not all measures will be developed for the 19 MOS. All project criterion measures will be developed for the following MOS.

1. 13B - Cannon Crewman
2. 64C - Motor Transport Operator
3. 71L - Administrative Specialist
4. 95B - Military Police
5. 11B - Infantryman
6. 19E - Tank Crewman
7. 05C - Radio Teletype Operator
8. 63B - Light Wheel Vehicle/Power Generation Mechanic
9. 91B - Medical Specialist

Measures of training success and service-wide performance will be developed for the following specialties.

1. 76Y - Supply Specialist
2. 94B - Food Service Specialist
3. 12B - Combat Engineer

4. 16S - MANPADS Crewman

5. 55B - Ammunition Specialist

6. 76W - Petroleum Supply Specialist

7. 54E - Chemical Operations Specialist

8. 67N - Utility Helicopter Repairman

9. 51B - Carpentry/Masonry Specialist

10. 27E - TOW/Dragon Repairman

RESEARCH DESIGN

Types of Measures to be Developed and Descriptions

Hands-on performance measures, job knowledge tests, and performance rating scales will be developed for training success, service-wide performance, and MOS-specific performance for the MOS listed above. A more complete description of these measures appears in the FY83 Annual Report (1983) and in the 2nd Annual Report to Congress (1983).

Measures Construction

Construction of the different measures will be based on job and task analysis data. A multi-method job analysis approach will be employed in which requirements are determined using existing Army job inventory procedures and by applying one or more judgmental approaches. For example, the critical incident technique will be used to develop performance rating scales, and the Army Occupational Survey job inventory approach will help identify important MOS-specific tasks. The accuracy, completeness and appropriateness of the task and job information obtained will be assessed in the development of the criterion measures. A detailed description of the task analytic and measures construction strategies appears in the FY83 Annual Report (1983) and Second Annual Report to Congress (1983).

Pretesting Strategy

The completed hands-on test package will be pilot tested with representative scorers and soldiers. The purpose of the field testing is to assure that the test can be administered as designed in a field environment and to determine scorer reliability. Subjective data on acceptability and feasibility will also be collected from scorers and examinees. The pretesting strategies are described in greater detail elsewhere.

Sampling Approach

The sampling plan specifies which MOS will be examined from the universe of possible MOS and details sample sizes for first-term enlisted personnel within each MOS. The selection of the 19 MOS was described previously. Detailed information regarding more extensive data collections planned for these and other measures is included in the research plan for "Improving the Selection, Classification and Utilization of Army Enlisted Personnel" (1983).

Data Collection Procedures

Data collection will begin with a briefing of local military commanders, examination of the test sites, equipment, and supplies, training of test administrators and scorers, and a dry run of testing procedures. Test site managers will be appointed to supervise the actual data collections and will be responsible for controlling the quality and flow of data from the testing.

Analyses to be Conducted

Statistical analyses will focus on reliability, validity and test fairness issues for the training, job, and service-wide performance measures. Reliability will be assessed using (1) test-retest procedures, (2) variance partitioning procedures to estimate generalizability of test results over variables such as task type, scorer, test station, etc., and (3) interrater agreement estimates. Validation of selection and classification measures will focus on the content, construct, and concurrent validity of the battery with respect to training, job, and service-wide performance. Analysis-of-variance techniques will be used to examine predictors for possible sex and race/ethnicity subgroup bias. Project analyses are discussed in full in the FY83 Annual Report (1983).

CURRENT STATUS AND ACCOMPLISHMENTS

A complete report of accomplishments and current status of the Army's research to link enlistment standards to job performance has been published by the Army Research Institute (1983) and is available for distribution. What follows is a summary of accomplishments in the area of job performance measurement during CY84.

Performance Measure Development

Job knowledge tests, hands-on tests, job performance and Army-wide rating instruments were developed for MOS 13B, 64C, 71L, and 95B. A report was prepared describing the rationale for and procedures followed in the analysis and selection of relevant job

tasks for the hands-on and job-specific knowledge tests for these MOS (HumRRO & AIR, 1984). The report presents recommendations for adoption in the analysis and selection of job tasks for other MOS.

Data Collection

Field Test. The criterion measurement instruments for MOS 13B, 64C, 71L, and 95B were field tested at three FORSCOM installations and two USAREUR bases. The field test MOS sample size by site data is shown in Table 1. The data are currently being analyzed to inform decisions about the FY85 concurrent validation and criterion development and field testing in other MOS.

Table 1

Field Test Sample Size by Site and MOS for 71L, 95B, 64C, and 13B

Site	71L	95B	64C	13B
Ft. Polk	60	42		
Ft. Hood	48	42		
Ft. Riley	21	29		
Mannheim			155	
Herzobase				150
Total	129	113	155	150

Longitudinal Database. The longitudinal database plan was completed (Wise, Wang, and Rossmeissl, 1983). The database supports the development and validation of new predictors and new measures of Army performance. The report describes the contents of the database, editing procedures, storage and access procedures, documentation and dissemination procedures, and security procedures. Data from the preliminary predictor battery administrations, existing training data, EMF quarterly data, and combined criterion field test data were edited and entered.

Analyses Conducted

Osborn and Hoffman (1984) discussed the relative cost effectiveness of hands-on and knowledge measures. Since cost-effective coverage of the task domain is a concern, it is important to use hands-on measures only where knowledge tests will not do. Thus, for many tasks in the four MOS above, both types of measures were constructed for comparison in a field test. These comparisons, made for various types of tasks, will enable an evaluation of the effectiveness of knowledge measures as surrogates for hands-on tests. The paper estimated the relative cost of hands-on and

knowledge tests, then examined measures of effectiveness, and, finally, discussed how these may effect the content validity of proficiency tests.

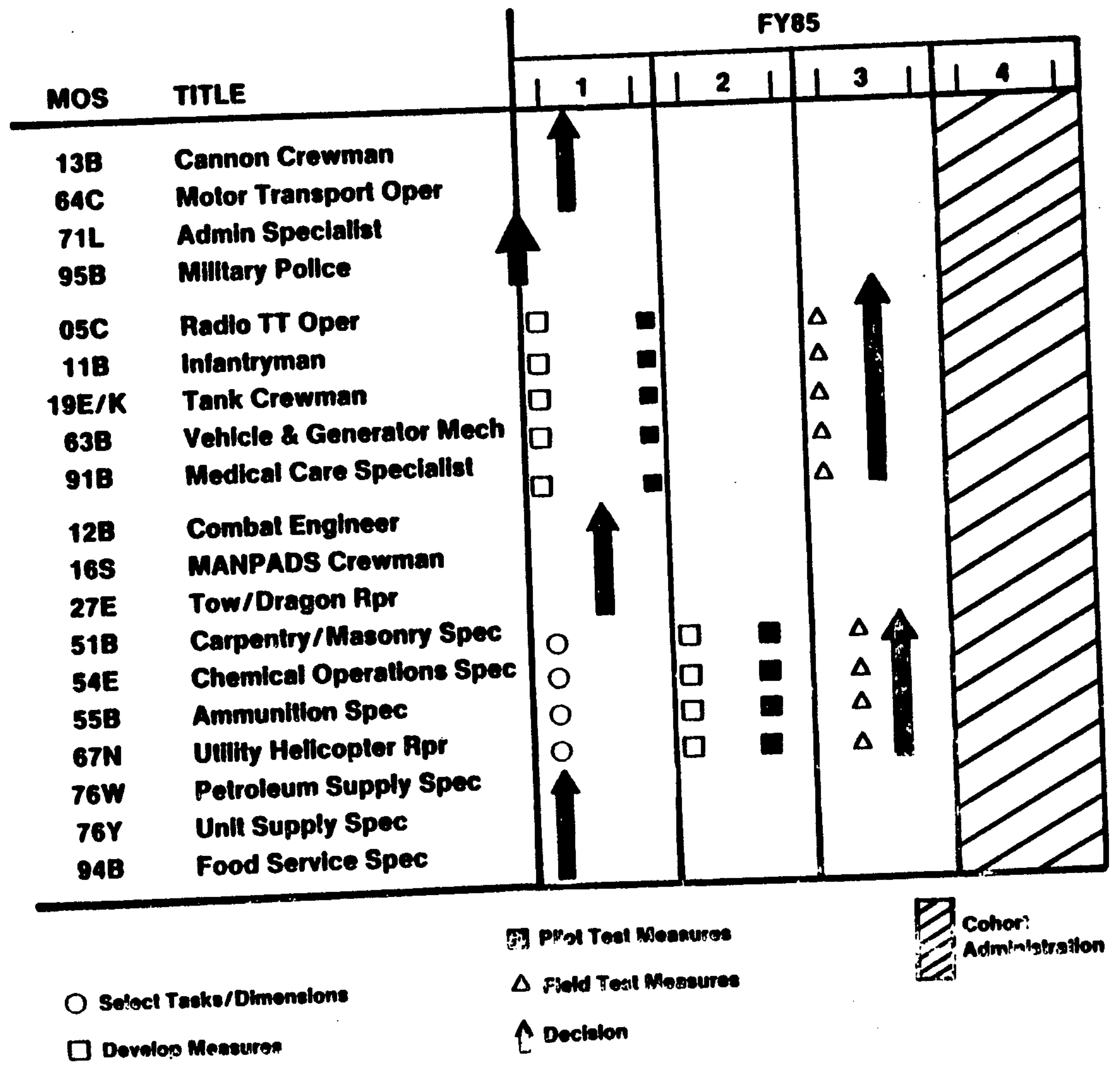
Borman, White, and Gast (1984) examined supervisor and peer ratings of first-term Army enlisted personnel as a function of several factors that potentially influence ratings. They considered such factors as 1) component job performance, 2) "good soldier" factors, 3) interpersonal/relationship factors, and, 4) job knowledge and skill factors. Peer and supervisor ratings were provided for 60 71L and 42 95B personnel. Correlations between overall job performance ratings and ratings on each of the factors identified as a potential influence on ratings were examined. Results suggested that ratings of overall job performance reflect more attention paid to individuals' performance on the job than to their standing on factors less directly relevant to performance.

Olson, Borman, Roberson, and Rose (1984) presented preliminary results of research to identify relationships between a soldiers' work environment and his/her performance. They discussed the development of an Army Work Environment Questionnaire (AWEQ) and some preliminary results from administering the questionnaire to 60 71L and 42 95B soldiers. The purposes of the research are: 1) to apply the critical incident methodology to identify environmental and situational influences that impact on job performance; 2) to develop questionnaire items which assess these positive and negative environmental factors encountered by soldiers during their first-tour of duty; and 3) to examine preliminary relationships between these environmental factors and Army-wide ratings (i.e., supervisory and peer) of overall soldier effectiveness. Riegelhaupt, Harris, and Sadacca (1984) described steps taken: 1) to determine which administrative indices in Army personnel records (reports of AWOL, Articles 15, Certificates of Commendation) have sufficient variance and acceptable base rates to warrant consideration in forming criteria and in-service predictors of soldier effectiveness; 2) to combine these indices within a model of soldier effectiveness; and, 3) to identify from which archival sources it is most feasible to obtain them.

MILESTONE SCHEDULE

The criterion instruments for each MOS are being developed and tried out following the FY85 milestone schedule shown. Instruments for some of the MOS were developed and tried out during FY84. These instruments are now being refined and a decision made on their content for administration during the FY85 concurrent validation. These decision points for all the MOS are shown on the milestone schedule.

FY85 Milestone Schedule



FUNDING/RESOURCES

Category	FY 1985	FY 1986	FY 1987	FY 1988
6.2	1.0M	1.0M	1.0M	1.0M
6.3	5.2M	4.6M	3.0M	3.0M
6.4	78K	80K	86K	89K

REFERENCES

Borman, W. C., White, L. A., and Gast, I. F. (1984, August). Fac
tors relating to peer and supervisor ratings of job perform-
ance. Paper presented at the annual convention of the American
Psychological Association, Toronto, Ontario, Canada.

Human Resources Research Organization (HumRRO) and American Insti-
tutes for Research (ARI). (1984). Selecting job tasks for
criterion tests of MOS proficiency. Research Report (in
press). Alexandria, VA: U. S. Army Research Institute for the
Behavioral and Social Sciences.

Human Resources Research Organization (HumRRO) and American Insti-
tutes for Research (ARI), Personnel Divisions Research Insti-
tute (PDRI), and Army Research Institute (ARI) (1983).
Improving the Selection Classification and Utilization of Army
Enlisted Personnel: Annual Report. US Army Research Institute
for the Behavioral and Social Sciences: Alexandria, Virginia.

Human Resources Research Organization (HumRRO) and American Insti-
tutes for Research (ARI), Personnel Divisions Research Insti-
tute (PDRI), and Army Research Institute (ARI) (1983).
Improving the Selection Classification and Utilization of Army
Enlisted Personnel: Project A Research Plan. US Army Research
Institute for the Behavioral and Social Sciences: Alexandria,
Virginia.

Office of the Assistant Secretary of Defense (Manpower, Reserve
Affairs, and Logistics) (1983). A Report to the House Commit-
tee on Appropriations: Second Annual Report to Congress on
Joint-Service Efforts to Link Standards for Enlistment to
On-the-Job Performance.

Olson, D. M., Borman, W. C., Roberson, L., Rose, S. R. (1984, Au
gust). Relationships between scales on an Army work environment
questionnaire and measures of performance. Paper presented at
the annual convention of the American Psychological Associa-
tion, Toronto, Ontario, Canada.

Osborn, W., and Hoffman, R. G. (1984, August). The cost-effec
tiveness of hands-on and knowledge measures. Paper presented
at the annual convention of the American Psychological Associa-
tion, Toronto, Ontario, Canada.

Riegelhaupt, B. J., Harris, C. D., and Sadacca, R. (1984). The
development of administrative measures as indicators of soldier
effectiveness. Research Product (Draft). Alexandria, VA:
Human Resources Research Organization (HumRRO).

Selection and Classification Working Paper 83-2

EXAMINATION OF CURRENT INFANTRY SCHOOL ACCESSIONS AND PROJECTION OF FUTURE NEEDS

Darlene M. Olson
Lawrence M. Hanser

United States Army Research Institute
for the
Behavioral and Social Sciences

5001 Eisenhower Avenue
Alexandria, Virginia 22333

Commercial: (202) 274-8275
Autovon 8 284-8275

DEPARTMENT OF THE ARMY
US ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22333

REPLY TO
ATTENTION OF

PERI-RS

18 APR 1983

SUBJECT: Study on Determination of Accession by Mental Aptitude

Assistant Commandant
US Army Infantry School
Ft. Benning, GA 31905

1. Last September you requested our assistance in determining whether the entrance requirements for MOSs 11B, 11C, 11H, and 11M should be modified for FY84-85. We assigned two scientists from our Manpower and Personnel Research Laboratory to assist you, Dr. Larry Hanser and Ms. Darlene Olson. I understand that they worked closely with your POC, CPT Flanagan, to accomplish this task. I am enclosing a copy of their final report. While I know that this information was presented to you in a formal briefing on the subject by CPT Flanagan, I am sure you will find it interesting reading.

2. We are glad to have been able to be of assistance to you and welcome the opportunity to do so again. Please feel free to call on us.

SIGNED

L. NEALE COSBY
COL, IN
Commander

1 Incl
as

CHIEF, S&C TA

PERI-RS

SUBJECT: Examination of current Infantry School accessions and projection of future needs.

TO: BG K.C. Leuer, Assistant Commandant, USAIS

FROM: U.S. Army Research Institute for the Behavioral and Social Sciences

1. References:

a. Letter, Infantry School, ATSH-I-V-ED, 10 Sep 82.

b. Indorsement, ARI, PERI-RP (10 Sep 82), subject: Distribution of Accession by Mental Aptitude.

2. In response to reference (1.a.) above, the following report is provided. The referenced letter requested that the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) review the current mental accession standards for CMF 11, and determine whether the mental ability requirements should be changed for MOSs 11B, 11C, 11H, or 11M during the FY 84-85 time period.

3. In order to clarify the scope of the requested Technical Advisory Service (TAS), two ARI psychologists met with Infantry School personnel on November 1-2, 1982 at Ft. Benning. As a result of these meetings it was agreed that the following task requirements would be investigated:

a. Examine whether available data provide any basis for adjusting the current minimum ASVAB Combat Composite Score (CO) or AFQT Score required for entry into an Infantry military occupational specialty (MOS).

b. Examine whether available data provide any basis for adjusting the current distribution of soldiers accessed into an Infantry MOS across AFQT mental categories.

c. Examine whether available data provide any basis for increasing the proportion of high school graduates accessed into Infantry MOS.

d. Examine whether any ability requirements can be identified to aid in assignment of a soldier to a particular MOS within the Infantry Career Management Field (CMF).

e. Examine the current procedures used to assign soldiers to a particular MOS within the Infantry CMF.

4. The purpose of this report is to summarize our findings with respect to the above listed task requirements.

5. ASVAB CO Score.

With respect to the task requirement in paragraph 3a above, examination of the data indicated that the Infantry accessed personnel of comparable ability to the Army in general, and substantially better in ability than Infantry personnel accessed during the period when ASVAB was misnormed. Evidence from examining the relationships between CO score and attrition rate in One-Station Unit Training (OSUT) and performance in the field as reflected by Skill Qualification Test (SQT) scores indicated that even a dramatic increase (from 85 to 105) in the minimum Infantry CO score would not significantly lower attrition in OSUT or substantially increase the probability of passing SQT.

6. Distribution of Accessions by Mental Categories.

Although our database did not contain adequate information to address all questions pertaining to the distribution of accessions by mental categories (paragraph 3b above), it did permit us to consider two such questions: (1) did the distribution of Infantry accessions by mental category in FY 1983 differ in any substantial way from the distribution of such accessions in FY 1982, and (2) how well are the Infantry accessions performing as soldiers? With respect to the first question, we found improvement in the distribution of Infantry accessions into the higher AFQT categories of I through IIIA from FY 1981 to FY 1983. In FY81, the proportion of CMF11 accessions in categories I through IIIA was approximately 44%. This figure rose to approximately 52% in FY82, and as of the end of the first quarter of FY83 had risen to approximately 60%. With respect to the second question, we found that at least 90% of the Infantry accessions in the mental category range of I-IIIA pass the SQT. The probability of passing the SQT for category IIIB is 83%; for category IVA is 77%; and for category IVB is 72%. Hence, the Infantry is currently accessing quality recruits who perform well as soldiers after training. These findings do not in our judgment provide any basis for adjusting the current distribution of Infantry accessions across AFQT mental categories I-IIIb, but suggest a reduction in category IV.

7. Education (High School vs. Non-High School Graduates).

In response to the task requirement in paragraph 3c above, OSUT attrition data revealed that non-high school graduates attrited at over twice the rate (15.5% vs. 6.2%) of

high school graduates. Also, Infantry first term enlisted personnel who are not high school graduates drop out of the Army, specifically in the I-IIIA and IIIB mental categories, at approximately twice the rate of high school graduates. While distributions of accessions by mental category showed increased numbers of enlistees in Categories I-IIIA, many of these were not high school graduates. Based on the differentially high attrition rate for Infantry accessions who are non-high school graduates and the large number of them in CMF 11, justification exists for the Infantry to attempt to access more high school graduates.

8. Ability Requirements for Infantry MOS

Data addressing the task requirement in paragraph 3d were collected from six company commanders and 3 non-commissioned officers assigned to training units in CMF11 at Fort Benning. Each of these subject matter experts responded to a computerized Job Assessment Survey. This survey is an experimental instrument developed under the direction of the U.S. Army Research Institute's Systems Laboratory. It was designed to aid in developing a profile of abilities required for successful performance in a specific job. For purposes of responding to your request it was used as an aid in determining which abilities are required for successful performance in the various CMF11 MOS. It was also used to aid in understanding how the ability requirements differ from one CMF11 MOS to another. The following represents a summary of our results:

a. Our data collection found that in general the abilities of memorization and spatial orientation were rated as highly required for successful performance in the CMF 11.

b. In the 11B MOS, in which the physical qualifications are among the most demanding in the Army, expert judgment (ratings) corroborated the importance of such physical requirements as stamina; the ability to perform flexible, speeded and coordinated muscular actions; and the ability to respond quickly to oral directions.

c. For the Indirect Fire Infantryman (11C MOS) the profile of average requirements showed that the ability to use basic arithmetic operations and the ability to make fine coördinated movements of the arms, hands, wrists and fingers were critical to successful performance.

d. In the 11H MOS, highly successful performance as a Heavy Anti-armor Weapons Crewman was linked to such perceptual-motor requirements as the ability to make precise controlled muscular movements, which would be necessary to adjust/position equipment; the ability to accurately distinguish similarities and differences when comparing visual patterns/objects; and the ability

to maintain arm-hand steadiness.

e. An examination of the requirements for the 11M MOS (Fighting Vehicle Infantryman) revealed that a complex group of cognitive and psychomotor abilities were important for successful performance. These requirements included the ability to evaluate and organize information in accordance with preplanned activities; the ability to allocate time among competing tasks; and the ability to selectively attend to relevant information.

Currently all CMF11 accessions are assigned to training as 11X. The assignment to training in a specific MOS such as 11H is made at the company level during BCT. The results of these analyses indicate that different skills may be required to perform successfully in these MOS. They also provide a preliminary indication of which skills are required for which MOS. At this time we do not have tests which adequately measure all these abilities. Such tests are being developed as part of our comprehensive soldier selection and classification research.

9. Assignment of Soldiers to Infantry MOS.

In response to the task requirement in paragraph 3e above, discussions with Infantry experts (SMEs) revealed that Army accession quotas, initial company platoon assignments, and pre-enlistment geographic locale preferences of individual soldiers determine to a great extent assignment to a specific Infantry MOS. These initial allocation factors, which restrict the available pool of soldiers, operate independently of soldier ability and performance in the assignment process. To the extent that different ability profiles are important in performance in a specific MOS, the present assignment system will result in errors in assignment and lowered overall performance. As noted earlier, assignment of individuals to specific MOS at the Infantry School is accomplished at the company level. Under these circumstances the commander has very little freedom for the reasons mentioned above. Perhaps more importantly, the individuals who make these assignment decisions have been provided with very little guidance on how to make them. One recommendation for improving these conditions would be to remove the decision to the battalion, brigade, or school level. This would provide more freedom in the assignment decisions by enlarging the eligible pool. At the same time arguments could be made to continue to make the decisions at the company level. After all, the company personnel have the closest contact with the individuals who are to be assigned, and perhaps know them best. One reasonable answer to this problem would be to provide broad guidelines to the company commanders to aid them in assigning individuals to specific MOS. Hence, although in terms of selection the Infantry as a whole is obtaining quality accessions, policies at both the DA and Army Training Center level which impact on classification or

assignment of soldiers to specific CMF11 MOS should be examined in order to better optimize the match between soldiers with certain abilities and specific requirements of an MOS.

Working Paper

RS-WP-87-03

Presentation of Sex-related Differences in Spatial Abilities and Performance of Army Soldiers for the Joint Services Selection & Classification Working Group

Darlene M. Olson and Elizabeth P. Smith
SELECTION AND CLASSIFICATION TECHNICAL AREA

May 1987

Martha L. Teplitzky

Reviewed By:
Martha Teplitsky
Personnel Utilization Technical Area

Clessen J. Martin (for)

Approved By:
Lawrence M. Hanser
Chief, Selection and Classification Technical Area

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

Spatial abilities influence effective adjustment to the physical environment and impact on how individuals perceive and interact with their environments. Research has shown that spatial abilities play crucial roles in academic performance, vocational choice, and job performance (e.g., Ghiselli, 1966; Fennema & Sherman, 1977; McGee, 1979). Historically, research on cognitive abilities has found that men tend to obtain significantly higher mean scores than women on measures of spatial relations, visualization and orientation skills (Maccoby & Jackson, 1974).

During the last decade this general finding of sex-related differences in spatial aptitude has been questioned (Fairweather, 1980; Hyde, 1981; Caplan, MacPherson, & Tobin, 1985). First, it has been argued that spatial ability is not a well-defined construct in terms of either subabilities or psychological process variables (Lohman, 1979). Second, findings regarding which spatial ability factors show sex-related differences, at which developmental ages, and in relation to which spatial tasks/test items have yet to emerge consistently. Third, the advent of meta-analytic techniques (Glass, McGaw, & Smith, 1981; Hyde, 1981) for examining the effect size of mean differences across individual research studies suggests that the magnitude of gender differences for visual-spatial ability is small, with 5% or less of the variance explained by sex. While the existence and importance of sex-related differences in spatial abilities is debated, the practical significance of even small differences between the sexes in spatial skills in terms of opportunities, perceptions, and performance expectations should not be ignored.

This research was conducted as a component of Project A, a large selection and classification research program initiated by the Army. The purpose of Project A is to validate the Armed Services Vocational Aptitude

Battery (ASVAB), the military's entry-level cognitive selection battery, as well as other new spatial, perceptual, psychomotor, and temperament measures against job performance criteria. Specifically, the purposes of the research reported here are to: (1) investigate sex-related differences in mean scores on the Project A paper-and-pencil spatial tests, (2) examine correlations between spatial tests and ASVAB subtests, (3) compare the factor structures of the spatial tests and ASVAB subtests for men and women, and (4) examine the correlations between scores on the spatial predictors and Project A performance constructs for the sexes.

METHOD

Subjects. The sample contained 4926 (females = 839 and males = 4087) first-term Army enlisted personnel in 9 jobs representing clerical/administrative, operations, and skilled/technical military occupational specialties (MOS). These jobs were sampled at eleven continental United States and four European Army installations. Table 1 describes the sample by sex and Army job.

Insert Table 1 about here

Cognitive Ability Measures. The operational Armed Services Vocational Aptitude Battery (ASVAB) was used in this research as a measure of general cognitive ability. The ASVAB, a paper-and-pencil instruments contains the following 10 independently timed and scored subtests: (1) Arithmetic Reasoning, (2) Mechanical Comprehension, (3) Math Knowledge, (4) Numerical Operations, (5) Paragraph Comprehension, (6) Auto/Shop Information, (7) Word Knowledge, (8) General Science, (9) Electronics Information, and (10) Coding

Speed. The 30 item Arithmetic Reasoning subtest requires the ability to implement the arithmetic operations of addition, subtraction, multiplication, and division of integers and fractions. The Mechanical Comprehension subtest, which contains 25 items, assesses knowledge of the properties of materials and structures and how mechanical devices operate. The Mathematics Knowledge test (25 items) requires the ability to perform formal algebraic operations. The speed with which a respondent can perform simple arithmetic operations on primarily single digit whole numbers is assessed by the 50 item Numerical Operations subtest. The Paragraph Comprehension subtest presents respondents with a brief reading passage and asks questions about the content of the passage which require the abilities to infer and generalize from what is read. The 25-item, Auto/Shop Information subtest measures knowledge of automobile repairs and facts about metal/wood shops. The vocabulary of respondents is measured with the 35 item Word Knowledge subtest. The General Science subtest measures knowledge of the physical and biological sciences and the Electronics Information subtest assesses practical knowledge about electricity and electronics. The Coding Speed subtest, which draws heavily on short-term memory requires respondents to compare symbols (i.e., a common word with a four-digit number) and transfer the information from the test booklet to the answer sheet as rapidly as possible.

The Project A spatial battery used in this research contains 6 timed paper-and pencil tests, with a total of 220 items and a time limit of 63 minutes. A 30-item Spatial (Figural) Reasoning test was used to measure soldiers' ability to determine the principles which govern the relationships among a group of abstract figures. This test is expected to assist in prediction of performance in such areas as identification and detection of

targets, troubleshooting and report work, and analysis of intelligence data. The visualization measures consist of (1) a 90-item Object Rotation test, that assesses the mental rotation of two-dimensional figures, (2) a 24-item Maze test, that measures spatial scanning or the ability to identify a specific pattern or configuration in a complex visual field, and (3) Assembling Objects, a 32-item measure of the ability to visualize assembled objects from their component parts. According to Peterson (1986) these visualization tests were expected to predict success in map drawing and usage in field operations related to mechanics and construction activities. Two orientation tests were also administered: A 24-item Orientation test, which assesses soldiers' ability to maintain their perspective or bearing with respect to some object that has been rotated, and the Map test, which measures the ability to orient oneself using a map and knowledge of compass directions. The spatial orientation tests were developed to predict success in activities which involve maintaining physical position relative to environmental landmarks or location under frequent directional changes (Peterson, 1986). The spatial tests are very reliable with split-half reliabilities corrected for length using the Spearman-Brown formula ranging from .87 (Reasoning) to .99 (Object Rotation).

Performance Measures. Criterion development work was conducted by the Project A contractors and included construction of the following measures: (1) Army-wide rating scales relevant for evaluating soldiers in any first-tour Army job, (2) job-specific rating scales, (3) hands-on task proficiency measures, and (4) job knowledge tests. The Army-wide rating scales were developed using a variant of the behaviorally-anchored rating scale methodology, and emphasize performance dimensions relevant to any MOS (e.g.,

maintaining equipment). The job-specific scales, which were also 7-point behavior summary scales, focus on narrow performance areas relevant to a designated job (e.g., transporting personnel for the motor transport operator job). The hands-on tests consisted of 15 MOS-specific tasks. Hands-on scores were computed for each soldier by averaging the proportions passed across the tasks tested. Multiple choice tests were developed to assess job knowledge relevant to import and representative tasks in an MOS. A total job knowledge score for each research participant was derived as a percentage of the number of items answered correctly. Factor-analysis of the performance ratings resulted in an interpretable solution with three factors: (1) Effort and Leadership, (2) Personal Discipline, and (3) Physical Fitness and Military Bearing (Campbell, Hanser, Wise, 1986). These represent the "will do" or motivational aspects of the performance domain. In addition, a General Soldiering Proficiency factor and a MOS-specific Core Technical Proficiency factor, which are based on factor analysis of written job knowledge and hands-on tasks, emerged as representative of the "can do" or more job/task specific components of performance. Hence, the interactive research process pursued in Project A has resulted in a predictor and criterion space that are multidimensional.

Specifically, then, the criterion domain as defined in Project A contains five performance constructs:

1) The Core Technical Proficiency construct represents the proficiency with which the soldier performs the tasks which are "central" to his/her MOS. It refers to how well the individual can execute the core technical tasks the job requires, given the motivation to do so.

2) The General Soldiering Proficiency construct represents the degree

to which the soldier can perform a variety of Army-wide general soldiering tasks, including use of basic weapons, NBC procedures, first aid, basic field techniques, etc.

3) The Effort and Leadership performance construct reflects the extent to which the soldier exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates the willingness to exercise leadership and support toward peers.

4) The Personal Discipline Construct reflects the degree to which the soldier adheres to Army Regulations and traditions, exercises personal self control, demonstrates constraint in daily behavior and does not cause disciplinary problems.

5) The Physical Fitness and Military Bearing performance construct represents the degree to which the individual maintains an appropriate military appearance and remains in good physical condition.

RESULTS

Descriptive Statistics: Table 2 presents the means, standard deviations, t-tests and effect sizes on the Project A spatial tests for total sample and by sex group.

Insert Table 2 about here

For the total sample, multivariate analysis indicated that there is a significant overall sex-related difference on the spatial tests ($T^2 = 279.30$, $p = .0001$). Univariate t-statistics show that men have significantly higher mean scores than women on all the spatial measures except Assembling Objects ($t = -1.08$, $p = .28$) and Spatial Reasoning ($t = -4.85$, $p = .0001$). The

magnitude of the significant sex-related differences on the spatial tests is on the average, less than 3/4 of a standard deviation. Further, the magnitude of the effect sizes were all small (based on Cohen's [1977] interpretations), indicating that only minimal mean differences exist between men and women on the Project A spatial tests.

Correlational Analyses. Tables 3 and 4 present the intercorrelation matrices for the Project A spatial tests and the Armed Forces Qualifications Test (AFQT) by total sample and by sex group. In the total sample, AFQT was

Insert Tables 3 and 4 about here

most highly correlated with the Map test and least correlated with the Object Rotation test. This finding was also observed for males and females separately. Intercorrelations among the spatial tests show that the orientation measures tend to be more highly related to each other than to the visualization tests. This finding lends some support for the convergent and discriminant validity of the separate measures.

Table 5 displays the correlations between the ASVAB subtests and the Project A spatial tests for the total sample. Generally the strongest

Insert Table 5 about here

relationships to the spatial measures are noted between the ASVAB subtests Arithmetic Reasoning, Mechanical Comprehension, and Mathematics Knowledge. This is particularly true for the Map and Spatial Reasoning Tests. Further, in general, correlations tend to be higher between the Map test and all ASVAB

subtests except Numerical Operations than for the other spatial tests. In comparison, the weakest correlations were found between the speeded ASVAB subtests (Numerical Operations and Coding Speed) and the spatial measures.

Correlations between the ASVAB subtests and the Project A spatial tests for males and females are displayed in Table 6. In general, slightly

Insert Table 6 about here

stronger relationships between scores on the ASVAB subtests and the spatial tests were observed for males than for females, probably due to greater variability among scores for males. As previously noted for the total sample, the largest significant correlations for both sexes were found between the spatial tests (e.g., Map and Spatial Reasoning) and the quantitative (e.g., Arithmetic Reasoning and Math Knowledge) and mechanical subtests of ASVAB.

Factor Analysis. Although overall mean scores on the spatial tests differ by sex, some commonalities in the structures were found for males and females in this research. Principal components factor analyses with varimax rotation, requesting a four-factor solution, were run for the total sample and for each sex group. These factor solutions are presented in Tables 7 and 8. For the total sample, Table 7 shows a strong spatial factor (Factor 1),

Insert Tables 7 and 8 about here

with all the Project A spatial tests loading at .60 or higher. Both quantitative and verbal ASVAB subtests load on Factor 2. Factor 3 is technical in

content, with strong loadings for Mechanical Comprehension, Auto/Shop, and Electronic Information. The speeded ASVAB subtests, Numerical Operations and Coding speed make up Factor 4.

Table 8 shows the separate factor structures by sex group. For females, Mechanical Comprehension loads more strongly with the Project A Spatial tests. For males' Mechanical Comprehension loads higher with the verbal and technical subtests from ASVAB than with the spatial tests. Also, for females, the spatial tests all load highest on the same factor with Mechanical Comprehension, but the Map and Spatial Reasoning Tests also load with the quantitative ASVAB subtests. In contrast, for males, the spatial tests tend to load on two separate factors. Specifically, Map, Orientation, and Spatial Reasoning load with the quantitative ASVAB subtests while Object Rotation and Maze form their own factor on which Mechanical Comprehension loads to a moderate degree. Assemble Objects loads equally high on both of these factors.

Correlations between Project A spatial tests and the five criterion constructs are presented for the total group in Table 9 and by sex group in Table 10. For the total group, correlations between the "can do" constructs (General Soldiering Proficiency and Core Technical Proficiency) are higher (range = .18 to .38) than for the "will do" constructs (Effort and Leadership, Personal Discipline and Military Bearing and Physical Fitness (range = .01 to .12), as would be expected. This pattern is the same for the two sex groups, but the range of correlations is higher for males.

Insert Tables 9 and 10 about here

CONCLUSION

In summary, as previous research in spatial abilities has found, the findings from this research provide support for significant sex-related differences on spatial abilities tests in favor of males. However, we also found a significant spatial difference in favor of females on Reasoning and no difference on Assemble Objects, a test of spatial visualization. Although significant mean differences do exist, the magnitudes of the differences and their actual effect sizes are not even moderate. Further, higher correlations among the Project A spatial tests themselves and in relation to AFQT and to the individual ASVAB subtests are found for males, but these differences are small also. In addition, patterns of intercorrelations for the two sex groups appear more similar than different. This is also true of the relationships between the spatial measures and the Project A criterion constructs. There are somewhat different factor structures for the two sex groups. These differences are not clear-cut, however, given that some tests and subtests have moderate to high loadings on more than one factor.

A general conclusion from these findings is that differences as a function of sex of the soldier do exist with the Project A spatial tests, but that these differences are not great. It remains to be seen, however, whether the differences are such that selection based on the spatial tests may be biased in favor of one sex group over the other. This important question must be answered. Analyses to address this issue are presently underway.

References

Campbell, J. P., Hanser, L. M., & Wise, L. (1986, November). Development of a model of Project A criterion space. Paper presented at the 28th Annual Convention of the Military Testing Association, Mystic, CN.

Caplan, P., MacPherson, G. & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. American Psychologist, 40, 786-799.

Cohen, J. (1977). Statistical power analysis for the behavioral science. New York: Academic Press.

Fairweather, H. (1980). Sex differences: Still being dressed in the Emperor's new clothes. Behavioral and Brain Sciences, 3. 234-235.

Fennema, E. & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. American Educational Research Journal, 14, 51-71.

Ghiselli, E. (1966). Validity of occupational aptitude tests. New York: Wiley.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA.: Sage Publications.

Hyde, J.S. (1981). How large are cognitive gender differences: A meta-analysis using omega-square and d. American Psychologist, 36, 892-901.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature. (Technical Report No. 8, Aptitude Research Project.) Stanford University: School of Education.

Maccoby, E. E. & Jacklin, C. N. (1974). Psychology of sex differences. Stanford, CA.: Stanford University Press.

McGee, M. (1979). Human spatial abilities: Sources of sex differences. New York: Praeger.

Peterson, N. G. (Ed.) (1986). Development and field test of the Trial Battery for Project A. (ARI Technical Report No.____). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Table 1

Description of the Sample

ARMY MOS	TOTAL	MALES	FEMALES
SINGLE CHANNEL RADIO OPERATOR (31C)	351	300	51
LT. WHEEL VEHICLE MECHANIC (63B)	612	571	41
MOTOR TRANSPORT OPERATOR (64C)	655	594	61
ADMINISTRATIVE SPECIALIST (71L)	494	221	273
PETROLEUM SUPPLY SPECIALIST (76W)	472	434	38
UNIT SUPPLY SPECIALIST 76Y)	609	513	96
MEDICAL SPECIALIST (91A)	481	357	124
FOOD SERVICE SPECIALIST (94B)	575	471	104
MILITARY POLICE (95B)	677	626	51
TOTALS	4926	4087	839

Table 2

Means and Standards for Total Sample, Females, and Males on the Project A Spatial Tests and Effect Sizes of Mean Differences by Sex.

SPATIAL TESTS	TOTAL SAMPLE			FEMALES			MALES					
	N	M	SD	N	M	SD	N	M	SD	EFFECT SIZE	t^{a}	p
ASSEMBLE OBJECTS	4924	23.26	6.68	839	23.47	6.29	4085	23.22	6.76	-.04	-1.08	NS
MAP	4919	7.41	5.41	837	6.33	4.91	4082	7.63	5.48	.24	6.85	.0001
MAZE	4920	16.28	4.76	837	15.11	4.83	4083	16.51	4.71	.29	7.61	.0001
OBJECT ROTATION	4923	61.93	18.93	839	57.22	18.39	4084	62.89	18.90	.30	8.08	.0001
ORIENTATION	4920	10.91	6.09	839	9.68	5.39	4081	11.17	6.19	.24	7.10	.0001
SPATIAL REASONING	4916	19.18	5.58	838	19.93	4.80	4078	19.03	5.71	-.16	-4.85	.0001

[a] (-) Sign indicates difference in favor of females

Table 3

Intercorrelations and Reliabilities of the Project A Spatial Tests and Armed Forces Qualifications Test (AFQT) for the Total Sample

TOTAL SAMPLE	1	2	3	4	5	6	r^a
1. ASSEMBLE OBJECTS							.91
2. MAP	.49						.90
3. MAZE	.50	.43					.96
4. OBJECT ROTATION	.40	.38	.49				.99
5. ORIENTATION	.46	.52	.40	.35			.89
6. SPATIAL REASONING	.56	.51	.43	.37	.48		.87
7. AFQT	.36	.54	.29	.27	.41	.50	

[a] Split-half reliability estimates using Spearman-Brown correction.

Table 4

Intercorrelations of the Project A Spatial Tests and Armed Forces Qualifications Test (AFQT) for Males and Females

TOTAL SAMPLE	1	2	3	4	5	6	7
1. ASSEMBLE OBJECTS		.50	.50	.40	.46	.55	.36
2. MAP	.44		.42	.38	.52	.53	.56
3. MAZE	.50	.40		.47	.39	.45	.30
4. OBJECT ROTATION	.42	.37	.54		.34	.38	.29
5. ORIENTATION	.46	.51	.39	.38		.49	.43
6. SPATIAL REASONING	.56	.45	.43	.38	.47		.51
7. AFQT	.36	.51	.32	.24	.34	.45	

NOTE: rs females below diagonal; rs males above diagonal.

Table 5

Correlations Between ASVAB Subtests and Project A Spatial Tests for the Total Sample.

	AO	MP	MZ	OR	OT	RS
ARITHMETIC REASONING	.41	.53	.32	.30	.43	.52
MECHANICAL COMPREHENSION	.45	.54	.40	.39	.49	.45
MATH KNOWLEDGE	.40	.51	.29	.26	.41	.48
NUMERICAL OPERATIONS	.01	.04	.08	.04	-.01	.08
PARAGRAPH COMPREHENSION	.22	.35	.18	.17	.28	.33
AUTO/SHOP	.28	.39	.27	.28	.34	.25
WORD KNOWLEDGE	.24	.40	.15	.15	.30	.34
GENERAL SCIENCE	.33	.48	.28	.27	.38	.40
ELECTRONIC INFORMATION	.26	.37	.23	.24	.32	.27
CODING SPEED	.15	.15	.19	.13	.08	.17

Note. The Project A spatial tests are: AO = Assemble Objects, MP = Map, MZ = Maze, OR = Object Rotation, OT = Orientation, and RS = Spatial Reasoning.

Table 6

Correlations Between ASVAB Subtests and Project A Spatial Tests for Females and Males

ASVAB SUBTESTS	SPATIAL TESTS											
	AO		MP		MZ		OR		OT		RS	
	F	M	F	M	F	M	F	M	F	M	F	M
ARITHMETIC REASONING	.39	.41	.47	.54	.32	.33	.28	.31	.40	.43	.46	.54
MECHANICAL COMPREHENSION	.45	.47	.51	.54	.44	.38	.35	.39	.46	.48	.43	.50
MATH KNOWLEDGE	.41	.40	.48	.52	.32	.29	.23	.27	.40	.42	.44	.49
NUMERICAL OPERATIONS	-.02	.01	.04	.07	.07	.11	.03	.07	-.07	.02	.05	.07
PARAGRAPH COMPREHENSION	.24	.22	.33	.37	.17	.19	.13	.19	.23	.30	.29	.33
AUTO/SHOP	.21	.32	.32	.40	.17	.26	.18	.27	.22	.34	.25	.31
WORD KNOWLEDGE	.22	.25	.33	.42	.19	.16	.11	.17	.20	.32	.25	.35
GENERAL SCIENCE	.30	.34	.42	.48	.24	.28	.19	.27	.29	.39	.37	.42
ELECTRONIC INFORMATION	.18	.30	.29	.38	.14	.22	.12	.23	.18	.32	.15	.33
CODING SPEED	.07	.16	.11	.19	.20	.23	.15	.16	.06	.11	.07	.18

Note. The Project A Spatial tests are: AO = Assemble Objects, MP = Map, MZ = Maze, OR = Object Rotation, OT = Orientation, and RS = Spatial Reasoning. *Ns* for females ranged from 837 to 839; *Ns* for males ranged from 4078 to 4085.

Table 7

Factor Loadings of the ASVAB Subtests and the Project A Spatial Tests for the Total Sample

	TOTAL GROUP (N = 4924)			
	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
ASVAB				
ARITHMETIC REASONING	.43	.59	.19	.27
MECHANICAL COMPREHENSION	.47	.34	.61	-.09
MATH KNOWLEDGE	.40	.62	.09	.33
NUMERICAL OPERATIONS	.01	.00	-.16	.85
PARAGRAPH COMPREHENSION	.08	.76	.17	.04
AUTO/SHOP	.21	.20	.81	-.15
WORD KNOWLEDGE	.08	.82	.19	-.08
GENERAL SCIENCE	.22	.69	.42	-.04
ELECTRONIC INFORMATION	.13	.32	.79	-.04
CODING SPEED	.11	.08	-.02	.82
SPATIAL TESTS				
ASSEMBLE OBJECTS	.77	.19	.06	-.01
MAP	.60	.41	.23	.07
MAZE	.72	-.04	.23	.15
OBJECT ROTATION	.64	-.08	.33	.12
ORIENTATION	.65	.30	.13	-.06
SPATIAL REASONING	.70	.40	-.02	.05

Note: Principal components factor analysis with a varimax rotation.

Table 8

Factor Loadings of the ASVAB Subtests and the Project A Spatial Tests for Females and Males

	FEMALES (n = 839)				MALES (n = 4085)			
	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
ASVAB								
ARITHMETIC REASONING	.27	.23	.75	.08	.42	.66	.10	.27
MECHANICAL COMPREHENSION	.43	.52	.32	-.01	.59	.37	.43	-.11
MATH KNOWLEDGE	.25	.18	.78	.19	.37	.70	.02	.32
NUMERICAL OPERATION	-.21	-.07	.21	.79	-.14	.11	-.02	.84
PARAGRAPH COMPREHENSION	.63	.01	.36	-.02	.70	.30	-.06	.11
AUTO/SHOP	.66	.21	.05	-.10	.69	.01	.44	-.14
WORD KNOWLEDGE	.75	.04	.21	-.13	.75	.36	-.09	-.03
GENERAL SCIENCE	.70	.13	.36	-.03	.76	.36	.11	.00
ELECTRONIC INFORMATION	.69	.13	-.02	.10	.75	.06	.30	-.04
CODING SPEED	.07	.16	-.04	.83	.07	.04	.19	.83
SPATIAL TESTS								
ASSEMBLE OBJECTS	.10	.67	.38	-.07	.09	.56	.56	-.04
MAP	.33	.48	.45	.05	.32	.62	.35	.06
MAZE	.12	.78	.09	.16	.05	.30	.72	.15
OBJECT ROTATION	.08	.79	-.01	.11	.14	.16	.72	.12
ORIENTATION	.09	.58	.44	.13	.20	.62	.36	-.07
SPATIAL REASONING	.11	.54	.54	-.04	.19	.69	.37	.04

Note: Principal components factor analysis with a varimax rotation.

PROJECT A
ARMY RESEARCH INSTITUTE

Table 9

Correlations Between Project A Spatial Tests and Criterion Constructs for the Total Sample.

CRITERION CONSTRUCTS	SPATIAL TESTS					
	AO	MP	MZ	OR	OT	RS
GENERAL SOLDIERING PROFICIENCY	.37	.38	.29	.23	.32	.37
CORE TECHNICAL PROFICIENCY	.35	.35	.25	.18	.28	.35
EFFORT AND LEADERSHIP	.12	.12	.11	.11	.08	.09
PERSONAL DISCIPLINE	.09	.06	.02	.01	.04	.04
PHYSICAL FITNESS AND MILITARY BEARING	-.04	-.07	-.04	-.03	-.07	-.05

Note. Project A spatial tests are: AO = Assemble Objects, MP = Map, MZ = Maze, OR = Object Rotation, OT = Orientation, and RS = Spatial Reasoning. Correlations greater than .03 are significant at $p < .01$. N ranged from 4812 to 4913.

Table 10

Correlations Between Project A Spatial Tests and Criterion Constructs for Females and Males

CRITERION CONSTRUCTS	SPATIAL TESTS											
	AO		MP		MZ		OR		OT		RS	
	F	M	F	M	F	M	F	M	F	M	F	M
GENERAL SOLDIERING PROFICIENCY	.34	.38	.34	.39	.22	.30	.22	.23	.29	.33	.33	.38
CORE TECHNICAL PROFICIENCY	.31	.36	.32	.36	.22	.26	.16	.19	.29	.29	.33	.35
EFFORT AND LEADERSHIP	.09	.13	.08	.12	.04	.11	.03	.11	.04	.07	.03	.11
PERSONAL DISCIPLINE	.03	.10	.02	.07	-.02	.03	-.03	.03	.01	.03	-.03	.05
PHYSICAL FITNESS AND MILITARY BEARING	-.01	-.04	-.04	-.08	-.01	-.05	-.02	-.05	-.02	-.08	.00	-.05

Note. Project a spatial tests are: AO = Assemble Objects, MP = Map, MZ = Maze, OR = Object Rotation, OT = Orientation, and RS = Spatial Reasoning. For females, correlations greater than approximately .09 are significant at $p < .01$. For males, correlations greater than approximately .04 are significant at $p < .01$. N for females ranged from 827 to 839. N for males ranged from 3984 to 4071.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-18

Project A: Current Status

Darlene M. Olson
Clinton B. Walker
Frances C. Grafton
Selection and Classification Technical Area

March 1989

REVIEWED BY:

Leonard A. White
Research Psychologist
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

for Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Introduction

Beginning in 1979 military testing has been the focus of increasing attention, from both the Department of Defense and Congress. The reasons for this interest are varied and include:

1) The discovery in 1979 that the then operational selection test battery for military enlistment was misnormed.

2) The switch in reference groups for military tests from the 1944 mobilization population (all men, including officers, serving under arms during that year) to the 1980 youth population (all 17- to 23-year olds in 1980 including both males and females).

3) A Congressional mandate to link entry test scores directly to on-the-job performance rather than to traditional training performance.

4) Intense interest in combining computer technology with emerging developments in testing theory.

As an integrated response to the various concerns the U.S. Army Research Institute designed Project A, "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel." Project A had many objectives including: (1) validation of the existing enlisted selection and classification battery, the Armed Forces Vocational Aptitude Battery (ASVAB), (2) development and validation of new cognitive and non-cognitive measures that add to the ASVAB in predicting total soldier performance, and (3) development of a comprehensive set of performance measures including paper-and-pencil tests of school knowledge and job knowledge, hands-on tests of job performance, and behaviorally-anchored rating scales.

The Project A research design consists of three main stages of data collection and analysis in an iterative progression of development, testing, evaluation, and further refinement of the selection/classification instruments (predictors) and measures of job performance (criteria). In the first stage, individual level data from FY 81-82 were examined to explore relationships between soldiers' ASVAB scores and their later performance in training and on first-tour Skill Qualification Tests (SQT).

Prior to the second stage of the data collection, 19 Military Occupational Specialties (MOS) were selected as a representative sample of the Army's 250+ entry-level jobs. These MOS were selected based on 1) a clustering of MOS based on similarity of job content, 2) the fact that these MOS are representative of Army jobs and account for over 45% of the Army's accessions, and 3) the recommendations of General Officers representing various Army Commands at that time. Nine of the 19 MOS were then

selected to have both paper-and-pencil and hands-on tests of job performance in addition to the school knowledge tests and ratings administered in all of the MOS. The nine fully-tested MOS in Project A are: Infantryman (11B), Cannon Crewman (13B), Tank Crewman (19E), Radio Operator (31C), Light Wheel Vehicle Mechanic (63B), Administrative Specialist (71L), Motor Transport Operator (88M), Medical Care Specialist (91A), and Military Police (95B).

The second stage of the Project A research, the Concurrent Validation (CV), was conducted in FY 1985. In the CV, over 9400 soldiers in the 19 MOS were administered the new Project A predictor tests, which included measures of spatial abilities, temperament and vocational interest, as well as computerized tests of perceptual and psychomotor skills. These tests were designed to expand the predictor domain in terms of individual characteristics and attributes that might be important for selection in predicting more aspects of performance. The new predictors were intended to supplement those cognitive abilities assessed by ASVAB.

Concurrently, the CV soldiers, who had been in the Army 12 to 27 months, were administered a comprehensive set of job performance measures including supervisory and peer ratings, written school and job knowledge tests, and MOS-specific hands-on task proficiency measures.

One of the major scientific contributions of the Project A research to date is a comprehensive modeling of the job performance domain. The criterion development efforts in the project were driven by the idea that job performance is multi-dimensional and that performance is best measured through a variety of methods. Analyses of the criterion data collected during the CV resulted in five empirically derived dimensions that represent overall Army job performance. The five Project A job performance dimensions are:

1) MOS-Technical Knowledge and Skill. The proficiency with which a soldier performs the technical tasks that are critical and central to his/her MOS.

2) General Soldiering Proficiency: How well a soldier executes common soldiering tasks such as first aid and land navigation.

3) Effort and Leadership: The degree to which a soldier exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers.

4) Personal Discipline: The degree to which a soldier adheres to Army regulations, exhibits self-control and does not create disciplinary problems.

5) <u>Physical Fitness and Military Bearing:</u> The degree to which a soldier maintains an appropriate military appearance and remains in good physical condition.

The first two performance dimensions involve a soldier's ability to perform the technical requirements of his/her job and are often referred to as "can do" factors. The remaining three performance dimensions are more attitudinal or motivational in nature and are often called "will do" factors.

The third stage of the Project A research, known as the Longitudinal Validation (LV), started in 1986 with the administration of the battery of new Project A predictor tests to approximately 55,000 new recruits in 21 MOS. At the end of his/her Advanced Individual Training (AIT) or One Station Unit Training (OSUT), each soldier was administered the school knowledge test for his/her MOS, and a set of rating scales was collected from supervisors and peers. Approximately 11,000 of these soldiers were followed into their first-tour field assignment and have been administered a set of MOS-specific and Army-wide job performance measures during FY 88 and FY 89. In addition, about 1,000 of the soldiers from the CV sample who reenlisted have been assessed on a battery of performance measures appropriate for second-tour including rating scales, written job knowledge tests, hands-on measures, and leadership measures.

<u>Results and Products</u>

During the execution of the research plan, Project A has made many scientific contributions to the field of Industrial and Organizational psychology and so far has delivered four major products to the Army.

<u>ASVAB</u>. One product which resulted from the Concurrent Validation research effort involves improvements in the computation and use of ASVAB Aptitude Area (AA) Composites. These composites are used to determine eligibility for training in Army jobs. Beginning with the operational implementation of ASVAB 11/12/13/14 in October 1984 the formulae for the Clerical (CL) and the Surveillance/Communications (SC) composites were changed. These changes resulted in both better accuracy and improved fairness in the prediction of job performance for minorities in several MOS. In addition, a change in the computation of the Mechanical Maintenance (MM) composite and recommended changes in required aptitude areas for approximately 50 MOS are scheduled for implementation in the near future. Annual savings from the changes in composites have been estimated at $25 million.

The Enlisted Personnel Allocation System (EPAS) is a new, computerized assignment system that will more efficiently match applicants to jobs for which they are best qualified, maximizing performance and minimizing attrition. Thus, with significant improvements in classification composites (from Project A) and in

assignment (from EPAS) the Army can put the "right" person in the "right" job at the "right time. The potential annual savings to the Army from using ASVAB (including the improved composites) in the Enlisted Personnel Allocation System (EPAS) to predict performance is estimated to be more than $480 million annually.

Results from the Project A CV have demonstrated conclusively that ASVAB is an excellent test of the "can do" or more technical task-based requirements of Army jobs. The mean validity coefficient between scores on ASVAB and Core Technical performance is .57 across the 19 MOS in the Concurrent Validation. Since ASVAB was designed as a general cognitive test, it is not expected to be a good measure of more specialized spatial abilities, psychomotor skills, motivation, interests or leadership. As a means of supplementing ASVAB, the new Project A tests measure these characteristics very well.

Spatial, Psychomotor and Perceptual Tests. Currently, ARI is supporting implementation of Project A's spatial and computerized psychomotor and perceptual tests in USAREUR and at training posts in CONUS.

These implementations were inspired by ARI selection research on tank and anti-tank gunners. In 1986, Project A's new tests had a multiple R of .76 against accuracy on high-tech simulators of tank gunnery for 95 Armor officers at Fort Knox. Cognitive ability, which was tested as well, did not contribute to the prediction. In 1987, 300+ new recruits at Fort Benning took the same battery before training in anti-tank gunnery. Several of the new tests strongly predicted accuracy in firing the TOW (Tube-launched, Optically-tracked, Wire-guided) missile simulator.

Based on these results, in December 87, CG TRADOC ordered implementation of Project A tests at four posts that train on high-tech weapons systems. ARI installed 1- and 2-hand tracking, maze, and mental rotation tests at Forts Knox, Benning, and Bliss in February 88.

At the Infantry and Armor sites, the earlier results have been confirmed. For 1,065 TOW students to date, 41% of those passing a cut score on the predictors qualified as gunners in the minimum time, as opposed to 24% of those not tested on the psychomotor/spatial predictors and 12% of those scoring below the cut score. Those scoring above the cut score qualified higher as well as faster: 48% attained the upper two levels of accuracy as contrasted with 36% of those not tested and 31% of those below the cut score. In terms of validities, accuracy was predicted at .37 by the Project A tests, .29 by ASVAB GT, and .38 by the two together. In mid-88, these training data were confirmed in live fire at Fort Benning: for 60 students firing one live TOW each at a moving target at 6,000', P(hit) was 85% for those who met the cut on the Project A tests and 73% for those who did not.

At Fort Knox, Armor recruits (N=500+) scoring in the upper third on the new predictors early in training later had gunnery hit rates 16% higher than those in the lower third, exactly repeating the results in the 1986 research. The new predictors correlated .54 with speed/accuracy, compared with a .34 validity for the ASVAB GT composite. Combined, the validity was .55, reconfirming the importance of spatial and psychomotor skills in predicting gunnery performance. At both posts, the new results were so positive that the test scores were incorporated into the decision making process.

Validities at Fort Bliss were not significant, and testing there has been suspended. In contrast to tank and anti-tank weapons, which rely heavily on tracking, the air defense systems tested use fire-and-forget weapons against fleeting targets. Thus these systems rely more on skills like vigilance and target identification.

Completing the implementation requests from CG TRADOC, the Field Artillery School, Fort Sill, is starting to administer a broader battery of the new tests to recruits in meteorology, surveying, radar range-finding, and artillery spotting (MOS). Validation will be against performance at the end of training.

Currently, the active forces are transitioning to the Bradley Fighting Vehicle. In USAREUR, V Corps has elected to use a battery of spatial and psychomotor tests to inform their selection of Bradley gunners in the 3rd Armor Division and the 8th Infantry Division.

A need for improved selection has been found also at the Special Forces School, where attrition, due primarily to failures in land navigation, is costly. In a new pilot project, ARI has installed three spatial tests at Fort Bragg to identify good land navigators.

In review, early positive results are being replicated in the implementations; spatial/psychomotor abilities strongly predict differences in gunnery performance; and utilization is spreading. Budget reductions, however, threaten the survival of the implementation in the training base. In the future, the tests could have the greatest impact if they were administered before enlistment. In that case the Army Reserve and National Guard could be served as well, and initial person-job matching could be strengthened in many MOS by adding spatial and/or psychomotor abilities to the profile of aptitudes.

<u>Assessment of Background and Life Experiences (ABLE)</u>. The ABLE is a 30-minute, multiple choice, non-cognitive test designed to measure temperament, personal history, and adaptability. In the Project A research, ABLE was shown to improve significantly the prediction of the motivational components of performance. Scores on the ABLE are strongly related to 12-month attrition. Recruits who have very low scores on the ABLE Adjustment

composite have attrition rates that are two to three times higher than soldiers with high Adjustment scores.

The Dependability scale of ABLE predicts in-service disciplinary problems. For example, soldiers having low Dependability scores receive significantly more Articles 15 than those with high scores. Conversely, soldiers with high Dependability scores are more often viewed as having potential of becoming high performing NCOs.

In sum, the ABLE shows promise to augment the Services' capability to identify the most qualified applicants in all AFQT categories and educational levels by predicting the "will do" components of performance. Based on these encouraging results, the ABLE is to be included in a joint Service adaptability screening instrument that will become operational in 1990.

The Longitudinal Research Data Base. The fourth, and perhaps most enduring product from the Project A research effort comes from the Longitudinal Research Data Base (LRDB). The LRDB is a permanent storehouse of empirical information on Project A, unparalleled in its richness for addressing recurring Army concerns. These data are invaluable to the Army in the areas of accession policy, setting standards for enlistment and reenlistment, predicting attrition, linking school training to field performance, and linking characteristics at entry into the military to performance in first and second tours of duty.

Project A was initiated in 1982 and will be completed in 1989. While Project A has focused primarily on Army enlisted personnel in their first tour of service, a second research program, "Building and Retaining the Career Force," will focus on the Army's career force, those soldiers who reenlist and remain beyond one tour of duty. Soldiers who comprise the career force provide not only leadership and continuity to the Army as an institution, but are crucial in insuring that the Army meets its battlefield mission.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-90-01

Project A: A Summary

Darlene M. Olson
Clinton B. Walker
Michael G. Rumsey
Frances C. Grafton
Selection and Classification Technical Area

September 1990

REVIEWED BY:

DAVID W. WITTER
Assistant Director
Manpower and Personnel
Research Laboratory

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

ZITA M. SIMUTIS, Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

**U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600**

INTRODUCTION

Since 1979 military testing has been the focus of increasing attention, from both the Department of Defense and Congress. The reasons for this interest are varied and include:

1. The discovery in 1979 that the then operational selection test battery for military enlistment was misnormed.

2. The switch in reference groups for military tests from the 1944 mobilization population (all men, including officers, serving under arms during that year) to the 1980 youth population (all 17- to 23-year olds in 1980 including both males and females).

3. A Congressional mandate to link entry test scores directly to on-the-job performance rather than to traditional training performance.

As an integrated response to these various concerns the U.S. Army Research Institute designed a large-scale research program to improve the Army's personnel system. The first phase of this program, Project A, focused primarily on initial selection and classification. Its objectives included: (1) validation of the existing enlisted selection and classification battery, the Armed Services Vocational Aptitude Battery (ASVAB), (2) development and validation of new cognitive and non-cognitive measures that add to the ASVAB in predicting total soldier performance, and (3) development of a comprehensive set of performance measures including paper-and-pencil tests of school knowledge and job knowledge, hands-on tests of job performance, and behaviorally-anchored rating scales.

Nineteen Military Occupational Specialties (MOS) were selected as a representative sample of the Army's 250+ entry-level jobs. Nine of the 19 MOS were then selected to have both paper-and-pencil and hands-on tests of job performance in addition to the school knowledge tests and ratings administered in all of the MOS. The nine fully-tested MOS in Project A are: Infantryman (11B), Cannon Crewman (13B), Tank Crewman (19E/K), Radio Operator (31C), Light Wheel Vehicle Mechanic (63B), Administrative Specialist (71L), Motor Transport Operator (88M), Medical Care Specialist (91A), and Military Police (95B).

The first stage of the Project A research, the Concurrent Validation (CV), was conducted in FY 1985. In the CV, over 9400 soldiers in the 19 MOS were administered the new Project A predictor tests, which included measures of spatial abilities, temperament and vocational interest, as well as computerized tests of perceptual and psychomotor skills. These tests were

designed to expand the predictor domain in terms of individual characteristics and attributes that might be important for selection in predicting more aspects of performance. The new predictors were intended to supplement those cognitive abilities assessed by ASVAB.

Concurrently, the CV soldiers, who had been in the Army 12 to 27 months, were administered a set of job performance measures including supervisory and peer ratings, written school and job knowledge tests, and MOS-specific hands-on task proficiency measures.

One of the major contributions of Project A to date is a comprehensive description of the major components of job performance for the first tour soldier. Many measurement methods were used in this project to assess all aspects of performance. Analyses of the data collected during the CV resulted in the identification of five performance dimensions, as follows: The five Project A job performance dimensions are:

1. MOS-Technical Knowledge and Skill. The proficiency with which a soldier performs the technical tasks that are critical and central to his/her MOS.

2. General Soldiering Proficiency: How well a soldier executes common soldiering tasks such as first aid and land navigation.

3. Effort and Leadership: The degree to which a soldier exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership and support toward peers.

4. Personal Discipline: The degree to which a soldier adheres to Army regulations, exhibits self-control and does not create disciplinary problems.

5. Physical Fitness and Military Bearing: The degree to which a soldier maintains an appropriate military appearance and remains in good physical condition.

The first two performance dimensions involve a soldier's ability to perform the technical requirements of his/her job and are often referred to as "can do" factors. The remaining three performance dimensions are more attitudinal or motivational in nature and are often called "will do" factors.

The second stage of the Project A research, known as the Longitudinal Validation (LV), started in 1986 with the administration of the battery of new Project A predictor tests to approximately 50,000 new recruits in 21 MOS. At the end of his/her Advanced Individual Training (AIT) or One Station Unit Training (OSUT), each soldier was administered the school knowledge test for his/her MOS, and a set of rating scales was collected from supervisors and peers. Approximately 10,000 of these soldiers were followed into their first-tour field assignment and were administered a set of MOS-specific and Army-wide job performance measures during FY88 and FY89. In addition, about 1,000 soldiers were assessed on a preliminary battery of performance measures appropriate for second-tour including rating scales, written job knowledge tests, hands-on measures, and leadership measures.

Results and Products

Project A delivered four major products to the Army.

ASVAB. One product which resulted from the Concurrent Validation research effort involves improvements in the computation and use of ASVAB Aptitude Area (AA) Composites. These composites are used to determine eligibility for training in Army jobs. Beginning with the operational implementation of ASVAB 11/12/13/14 in October 1984 the formulae for the Clerical (CL) and the Surveillance/Communications (SC) composites were changed. These changes resulted in both better accuracy and improved fairness in the prediction of job performance for minorities in several MOS. In addition, a change in the computation of the Mechanical Maintenance (MM) composite and recommended changes in required aptitude areas for approximately 50 MOS are scheduled for implementation in the near future. Annual savings in reduced attrition and improved job performance from the changes in composites have been estimated at $25 million.

Results from the Project A CV have demonstrated conclusively that ASVAB is an excellent test of the "can do" or more technical task-based requirements of Army jobs. The mean validity coefficient between scores on ASVAB and Core Technical performance is .57 across the 19 MOS in the Concurrent Validation. Since ASVAB was designed as a general cognitive test, it is not expected to be a good measure of more specialized spatial abilities, psychomotor skills, motivation, interests or leadership. As a means of supplementing ASVAB, the new Project A tests measure these characteristics very well.

Spatial, Psychomotor and Perceptual Tests. For success in tube-launched, optically-tracked, wire-guided (TOW) missile and tank gunnery, two types of special skills are required: one is the ability to deal with the orientation, location, and shape of

objects (spatial ability); the other is eye-hand coordination (psychomotor ability).

Project A developed new tests of these special skills. Spatial abilities are tested on paper with mazes and mental rotation items (where the soldier imagines how something will look after it has been turned). Psychomotor ability is tested on a device resembling a video game. Using hand controls, a soldier tries to keep a crosshair on a moving target.

Late in 1987, GEN Maxwell R. Thurman, then Commanding General of TRADOC, reviewed the results of preliminary research on these tests and directed the tests be installed at selected posts where gunnery skills are trained. From February 1988 through August 1989, Forts Benning, Knox, and Bliss tested over 24,000 new soldiers. At the end of training, many of these soldiers were tested on live fire or simulator gunnery performance. Students in the 11H TOW Gunner course who had high gunnery aptitude (i.e., upper 40 percentiles) on the Project A tests qualified in fewer trials and were more accurate than those with low gunnery aptitude. For example, the high aptitude students had a hit rate in live-fire that was 11% higher than the other students. These findings were replicated on TOW gunners in USAREUR in 1989, the hit rates in live fire being 92% and 60% for high and low aptitude soldiers, respectively. Similar results emerged at Fort Knox: 19K Armor students with high gunnery aptitudes had hit rates 7% higher than the average, and they acquired targets faster. Currently, USAREUR units are evaluating the usefulness of these Project A tests for the selection of Bradley gunners.

Thus, spatial and psychomotor abilities have been found to strongly predict differences in gunnery performance. Currently, a battery of Project A tests is being administered at Forts Benning, Knox, and Sill as a step toward possible full-scale implementation in the Joint Service selection testing process.

Assessment of Background and Life Experiences (ABLE). The ABLE is a 30-minute, multiple choice, non-cognitive test designed to measure temperament, personal history, and adaptability. In the Project A research, ABLE was shown to improve significantly the prediction of the motivational components of performance. Scores on the ABLE are strongly related to 12-month attrition. Recruits who have very low scores on the ABLE Adjustment composite have attrition rates that are two to three times higher than soldiers with high Adjustment scores.

The Dependability scale of ABLE predicts in-service disciplinary problems. For example, soldiers having low Dependability scores receive significantly more Articles 15 than those with high scores. Moreover, soldiers with high Depen-

dability scores are more often viewed as having potential of becoming high performing NCOs.

In sum, the ABLE shows promise to augment the Services' capability to identify the most qualified applicants in all AFQT categories and educational levels by predicting the "will do" components of performance. Because of the encouraging results obtained thus far, the ABLE is now being examined as a potential tool in the selection process for both the Rangers and Special Forces. It has also incorporated into a joint Service adaptability screening instrument that is being considered for initial operational testing in 1991.

The Longitudinal Research Data Base. The fourth, and perhaps most enduring product from the Project A research effort comes from the Longitudinal Research Data Base (LRDB). The LRDB is a permanent storehouse of Project A data, unparalleled in its richness for addressing recurring Army concerns. These data are invaluable to the Army in the areas of accession policy, setting standards for enlistment and reenlistment, predicting attrition, linking school training to field performance, and linking characteristics at entry into the military to performance in the first tour of duty.

Follow-on Research to Project A

Project A was initiated in 1982 and was completed in March, 1990. By demonstrating that the principal current selection and classification tool, the ASVAB, was doing an effective job of predicting performance, it provided justification for continued use of this valuable tool. Further, it generated new tests and demonstrated that substantial improvements could be made to the current system by adding these tests.

What remains is to use the information collected during the Longitudinal Validation of Project A to determine how best to combine old and new tests in a revamped, improved, initial entry screening process. This is one of the objectives of Building the Career Force, the second phase of the program to improve the Army's personnel system. A second objective of this follow-on project will be to provide a foundation for improved reenlistment and promotion decisions. When Building the Career Force is complete, the Army will have the basis for building a system which integrates personnel actions from a soldier's initial entry into the Army through promotion to the rank of sergeant, with all actions guided toward the goal of maintaining maximum readiness and fighting effectiveness of the entire Army organization.

Working Paper

RS-WP-84-20

A SUMMARY OF THE DEVELOPMENT PROCEDURES FOR
JOB PERFORMANCE MEASUREMENT IN PROJECT A

William Osborne

July 1984

Reviewed by
John Campbell

Approved by
Newell K. Eaton
Chief
Selection and Classification
Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

A SUMMARY OF THE DEVELOPMENT PROCEDURES FOR

JOB PERFORMANCE MEASUREMENT IN PROJECT A

This paper summarizes procedures being used in Project A to develop hands-on performance measures for Army job specialties. Job and task analysis procedures are covered in four phases: defining the task domain, identifying task characteristics, selecting tasks for testing, and detailing task procedures. A section on criterion test development is also included.

Since job specific performance tests are to be tied to job tasks, job performance must first be broken down into tasks, a sample of those tasks selected, then each further analyzed into its elemental parts. While a task inventory approach to job analysis has some disadvantages in that important contextual factors and translation behaviors can be lost in the "seams" when a job is partitioned into tasks, the approach has a number of important advantages. Chief among them is the exceptional degree of control and standardization that can be achieved in testing performance task by task. Another is analytic convenience: judgments about mission relevance, frequency of performance, kinds of behavior, and other data important to achieving content validity can be obtained systematically and reliably on a task by task basis. A final advantage particular to this project is that the Army's occupational analysis system is based on a task inventory approach.

Defining the Task Domain

The first stage of job analysis entails identifying the population of entry-level (Skill Level 1) tasks for an Army job specialty (MOS).

The Army Occupational Survey Center (AOSC), working with the technical school responsible for job and task analysis, routinely develops task lists for the various job specialties. A survey form based on the task list is used to obtain data from job incumbents on, among other things, tasks performed and frequency of performance. The task content that is prescribed for this job by Army management is listed and described in some detail in the Soldier's Manual (SM) for an MOS. The Soldier's Manual descriptions and related survey data encompass the range of doctrine (SM) and practice (AOS) in job performance. The two sources of task data differ also in the level of task statement. The Soldier's Manual describes job content at a somewhat more general level than do the survey items. To consolidate the two sources, a five step procedure is followed.

The first step is to use the occupational survey data to identify AOS tasks or activities that are reported as actually performed by Skill Level 1 (SL1) incumbents. This is a matter of screening the frequency-of-performance data for all AOS activities to which SL1 incumbents responded, and retaining those with non-zero frequencies.

The next step is to map the SL1 survey items onto tasks prescribed in the SM. An AOS activity is placed under an SM task if the statement duplicates the SM task or is subsumed under the SM task as a step or variation in conditions. The intent is to identify SL1 SM tasks with which the AOS activity can be matched. If this can not be done, higher skill level tasks described in the SM are successively reviewed and the AOS questionnaire items matched where possible with those SM tasks. The objective is to match AOS statements of practice with the SM statements of doctrine where possible, even if doctrine does not specifically identify the activity as a

SL1 responsibility. All SL1 SM tasks are retained regardless of whether or not they had parallel AOS statements.

Since some AOS questionnaire items can not be matched with any SM task, or any subset of elements from an SM task, the third step is to edit the remaining AOS task statements (i.e., items) so that, while similar in format to the SM task statements, they are still a clear portrayal of additional task content not contained in the SM. In some cases an AOS statement becomes a task statement by itself; in other cases a new task statement is developed which appropriately subsumes several AOS statements. The objective is to write all task descriptions at the same level of generality.

The final step is to have the consolidated list of task statements reviewed by subject matter experts (SME) at the proponent school. Usually three senior NCO's review the list for an MOS. The review focuses on the correctness of the matching of survey items with SM task descriptions, the appropriateness of the task titles for the AOS-only tasks, and relevance of tasks in light of the latest changes in equipment or doctrine. The consolidation of SM and AOS results in an MOS task domain for Skill Level 1. The domain includes all SL1 tasks from the Soldier's Manual for a particular MOS and their supporting AOS statements, all higher skill level tasks with supporting AOS statements, and AOS-only tasks. Described in this way, the domain portrays in precise task descriptive terms a definition of the job-world that an SL1 incumbent will face.

But even with consolidation the domain typically tends to span 200 to 300 tasks which is too many to be managed feasibly by SMEs from whom ratings of task criticality, difficulty and similarity are to be obtained.

So the domain is subjected to further consolidation. First, highly similar tasks that differ only because of equipment variations are consolidated. Next, tasks are deleted that pertain to peripheral or restricted duty positions in the MOS. Also deleted are higher skill level tasks doctrinally designated as irrelevant to SL1 job performance. Then any AOS-only (i.e., non-doctrinal survey derived) tasks with atypically low frequency-of-performance indexes are dropped. Finally, any existing data on task criticality may be used to eliminate tasks judged by job experts to be unimportant.

Application of these final filters tends to narrow the domain to approximately 150 tasks.

Identifying Task Characteristics

All tasks in the domain are now viewed as candidates for testing. Since the objective is to select from the domain a sample of tasks that is content valid, it is necessary at this stage to collect data on task characteristics pertinent to content validity. For Project A these have been designated as task importance, difficulty, and similarity. Data on frequency of performance is, as mentioned, available from AOS.

Using the rating protocols shown as Attachment 1, data on these task characteristics are obtained from 15 job experts (SMEs) at the proponent school. For each set of ratings--importance, similarity and difficulty--an SME works with a deck of task cards, each card listing the title of a task and a brief description of the task scope. In separate sessions, SMEs provide the three types of judgments: each (1) ranks the tasks as to importance in carrying out a prescribed unit mission, (2) sorts the tasks into

groups on the basis of similar performance requirements, and (3) estimates the typical distribution of soldiers over five categories of task proficiency.

The importance index is the rank of the task averaged over judges. Similarity is indexed two ways: one is the number of judges who placed the task with every other task; the other is cluster membership, that is the cluster containing the task, which is identified through a factor analysis of the cross-products matrix derived from the judges' task similarity clusters.

These task variables, together with frequency, are used in the next stage of the analysis, that of task selection--the goal being to select for testing tasks that are important, frequently performed, of moderate difficulty, and that represent a range of content.

Task Selection

Data from SME ratings on task importance, difficulty and similarity are used along with the occupational survey (AOS) data on frequency of performance to select 30 tasks for testing. That number was judged to approach the outside limit that could be handled by the resources of the project. Job analysts on the project staff process the data following a four phase procedure.

In the first phase, analysts are given the task data derived from the job experts' ratings and asked to select 30 tasks to represent each MOS, so as to maximize the content validity of the task sample. No strict rules are provided to the analysts for making their selections.

The second phase employes the use of policy capturing to identify each analyst's general tendency in using the data they were provided. Then, each analyst's task selections for each MOS are regressed on the task characteristics data. The resulting regression equations identify the implicit policies used in selecting tasks. The equations are then applied to the task characteristics data to provide an estimate of the task selections each would have made if their selections were completely consistent with their general tendency, as represented by a linear model.

In the third phase, analysts are given their original task selections and the selections predicted by their regression-captured policies. Differences may be expected between the two lists of tasks. Some deviations are expected because an analyst may have used something other than a linear combination of the task data or may have used knowledge about the MOS tasks not represented in the task characteristics data provided. On the other hand, some deviations may be due simply to unsystematic treatment of the data. Therefore, the purpose of phase three is to have the analysts review and justify discrepancies between their observed and predicted selections. Discrepancies that resulted from unsystematic treatment of the data can be corrected, and the rationale for each intentional discrepancy can be identified.

The final phase of the procedure is a Delphi-type negotiation among analysts to converge their respective choices onto 30 tasks for each MOS. The choices and rationale produced in phase three for each analyst are formatted and distributed to all analysts. Analysts can then retain or adjust their selections based on the opinion of others. The revisions are collected, formatted and redistributed for the analysts to make further

revisions. This process is repeated until consensus is reached on 30 tasks for each MOS.

The sample of tasks selected for testing by the measurement specialists is then reviewed by a panel of job experts from the proponent school for relevance, completeness, and fairness.

In the final stage of task description for test development, detailed descriptions of the steps and procedures in task execution are assembled for each selected task. Existing descriptive data from Field Manuals, Technical Manuals, Soldier's Manuals, school lesson plans, and other task analysis research are used in detailing the procedures, conditions and standards for performing a task (see examples in Attachment 2). Accuracy and completeness of the descriptions are confirmed through additional job expert reviews. The purpose for detailing these aspects of task performance is to provide the "items" that will be scored either in a paper-and-pencil or hands-on testing mode.

To bridge the gap between the assembled task detailing and hands-on performance measures it is necessary to determine the test-task boundaries. To do this, the developer determines the administrative limitations at the test site. The primary limitation is usually time; however, equipment availability, environmental limitations, terrain, support troops and potential equipment and personal risks also affect how much of the task can be tested. These factors often dictate that the task be reconfigured into one that meets test feasibility requirements and the task description is modified accordingly. But, before test development begins, the developer must make a final check to confirm that the modified, feasible task still reflects the task characteristics that dictated its inclusion originally.

With the assembling of the task analytic material and a rough outline of the hands-on scope of the task, test development begins.

Selection of Test Mode

The composite MOS test to be administered to cohorts will consist of a mix of hands-on and knowledge measures, each used for the type of task to which it is best suited. If knowledge tests correlated high enough with hands-on tests across the range of task types, we could rely on knowledge tests alone, saving substantially in administrative time and resources. But at this stage we can only speculate about the kinds of tasks that can and cannot be tested validly in a knowledge mode. In Project A our intent is to examine through a field test the knowledge-hands-on relationship for a variety of tasks, deciding on the basis of the correlational evidence those tasks for which a knowledge test may be used and those for which it may not.

So, for the experimental field test, both kinds of tests are developed for half of the tasks in an MOS. (Hands-on tests for 15 tasks is all that administration time will support.) The characteristics of tasks for which hands-on tests are expected to be necessary are: (a) psychomotor skill, (b) critical time limits on performance, and (c) performance steps uncued by task sequence or task conditions (i.e, steps that cannot be tested in a multiple-choice knowledge test without cueing the examinee). A rating procedure focussing on these three task characteristics is used by three task analysts to evaluate each of the 30 tasks. Tasks with high ratings on all three dimensions are the first designated for hands-on test development, since we suspect these will be most likely used in cohort measurement. The hands-on set is filled out with tasks representing the remaining range of scale values.

Knowledge tests are developed for all 30 tasks in the sample.

Test Construction

Development of a hands-on test begins with the task descriptive data and proceeds through four stages. The scoring approach is determined first. Here a decision is made as to whether test scoring will be directed at the product or the process of task performance, or both. The question to be answered is, what must the scorer observe to evaluate a soldier's performance? If task performance results in a product, then scoring can focus on the measurable characteristics of that product. However, since the relationship of task process to task outcome is not always evident, some element of performance--safety precautions, for instance--must be observed and scored as they occur.

In the second stage of development these product and process characteristics are translated into a score sheet or list of observable items to be evaluated by the scorer. It should be noted that these scoreable items, or what we term performance measures, are not merely a transcription of steps from the task description. Some task elements may have to be broken down into smaller parts if scoring is to be done reliably. In other cases, a step may be eliminated altogether from the scoring protocol because it is neither overtly observable nor produces an observable result. Included as performance measures are steps or elements molecular enough to be scored reliably by a trained evaluator with or without scoring aids. Dimensions of overall task performance such as time and sequence are added to the scoring protocol where important.

The third stage in developing a hands-on test involves preparing detailed specifications for test administration, including the precise conditions (environment, equipment, terrain, etc.) for testing, instructions to examinees, and instructions to scorers.

Once the draft test package has again been reviewed by job experts, development moves to the fourth and final stage in which through a small scale tryout the test is evaluated for administrative feasibility, scorer reliability and acceptability to examinees. Final revisions are guided by the results of this trial run. Sample hands-on tests for two Artilleryman (MOS 13B) tasks are given in Attachment 3.

Development of the knowledge test begins with the same task descriptive data. But here, in contrast to hands-on measures which are limited to observable task elements, covert or mental steps in task performance can be included. Methodically linking test questions to task procedures is the key to valid tests of job knowledge. And the sequence of decisions and actions to be followed in that linkage hinges on the causes of failure to perform the task correctly. Each essential behavior within the task is analyzed, with the help of a job expert where necessary, for potential causes of failure: Is it because the soldier doesn't know <u>when</u> to perform a step? Is it because the soldier doesn't know <u>where</u> to perform it? Is it because the soldier doesn't know <u>what</u> the end result looks like? Is it because the soldier doesn't know <u>how</u> to execute the behavior?

For each likely cause of error, the <u>correct</u> location, or sequence, or product, or procedure is identified and then described in words or pictures. A question is then framed, and, finally, real-world response

alternatives (distractors) are selected to complete the test item. The important point is that by considering these four questions about each aspect of task performance, we can pinpoint both what is important to ask in a knowledge test of task performance, and how to ask it. This procedure normally produces from five to 15 questions, depending on task complexity, that tap the critical aspects of task performance. It also helps prevent test questions that so often are used merely because they are easy to ask. Knowledge tests for two Artilleryman (MOS 13B) tasks are shown at Attachment 4.

Empirical Evaluation

The task analysis and test development process are designed to provide content validity, via the consensus of subject matter experts, for a) the completeness and relevance of the task domain, b) the representativeness and relevance of the task sample, and c) the description of the steps and procedures that are entailed in the execution of each task.

There are also several steps in the empirical evaluation of the reliability and validity of the task performance measures.

1) Several small pilot samples are being used to obtain the initial reaction of test takers and test scorers, to detect extreme response frequencies, and to estimate inter-scorer agreement.

2) A major field test using 150 subjects per MOS will be used to determine the test-retest reliability of the hands-on performance measure and to estimate the pattern of relationship of the hands-on measures with the parallel paper-and-pencil measures and other criterion measures in total Project A criterion arrays.

Selection and Classification Technical Area Working Paper RS-WP-88-6

INCORPORATING NEW PREDICTORS INTO THE ACCELERATED CAT-ASVAB PROJECT (ACAP) SYSTEM

Randolph K. Park

May 1988

Approved for Distribution to Sponsor or Proponent.

REVIEWED BY:

Leonard A. White
Research Psychologist

APPROVED BY:

for **Lawrence M. Hanser**
Chief, Selection and
Classification Tech Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

INCORPORATING NEW PREDICTORS INTO THE ACCELERATED CAT-ASVAB PROJECT (ACAP) SYSTEM

CONTENTS

LIST OF APPENDICES

INCORPORATING NEW PREDICTORS INTO THE ACCELERATED CAT-ASVAB PROJECT (ACAP) SYSTEM

INTRODUCTION

A future computerized testing system must have the capability to administer both the computerized adaptive test (CAT) version of the Armed Services Vocational Aptitude Battery (ASVAB) and "future tests" (FTs). The FTs must follow the ASVAB to protect the equating of CAT to the original version on paper-and-pencil (PP). These FTs must either have good incremental validity on criteria that ASVAB does not predict, or be candidates to upgrade the content domains that ASVAB does test. FTs will eventually have equal status with the ASVAB subtests, and be used to create Service area composites.

This paper is a report of the issues involved in implementing new predictors on the Accelerated CAT-ASVAB Project (ACAP) system. First, the content domains being investigated by the Services will be briefly summarized. Finally, relevant issues on ACAP implementation will be identified.

BACKGROUND

The Department of Defense (DoD) is currently sponsoring a project to administer the CAT-ASVAB. The responsibilities of the CAT Project are being shared by the four Services. The Air Force was responsible for developing the item pools. The Marine Corps contributes to the technical quality of the project. The Navy is the lead laboratory for research and development. The Army is responsible for the eventual procurement and fielding of a CAT system for full-scale testing.

As originally conceived, the CAT Project consisted of specially designed testing computers capable of meeting a number of criteria. One of these criteria, expansion/flexibility, required that any CAT hardware have the expansion capability to accommodate new predictors that the Services are developing. For example, with regard to Army needs, the CAT hardware must be able to accommodate tests being developed by Project A (Peterson, 1987).

In late 1984, the Office of the Secretary of Defense (OSD) directed early implementation for a form of CAT. This realigned the CAT Project's focus to the current Accelerated CAT-ASVAB Project (ACAP). Implementing the CAT-ASVAB on a limited scale is a goal of the ACAP. ACAP uses off-the-shelf computers rather than specially designed testing computers for faster implementation. ACAP is a "lessons learned" project, and will provide information for the eventual full-scale development (FSD) of CAT-ASVAB for nationwide implementation.

In November 1985, the Future Testing Committee was created by the CAT-ASVAB Working Group to explore the potential of CAT to include non-ASVAB new predictors now being developed by the Services. This committee was charged with evaluating such tests for addition to the ACAP system. The committee's purposes are to make recommendations of promising selection and classification tests, and to identify and address issues related to implementing new predictors with ACAP.

NEW PREDICTORS

Future tests are important for selection and classification purposes. While CAT-ASVAB testing can achieve test scores that are at least as reliable as conventional PP test scores (Moreno et al., 1985), little benefit can be derived from changing the mode of administration. At the composite level, Divgi (1987) has shown that CAT reliabilities may show only slight improvements because paper-and-pencil (PP-ASVAB) composite reliabilities already exceed .90. Thus increased measurement precision from CAT allows little gain in reliabilities. Even less gain is expected if CAT test scores are equated to PP test scores. As for validity, CAT will unlikely surpass PP-ASVAB because the same criterion measures will be used.

The best prospect for incremental validity is in content domains not measured by the ASVAB. For example, the Army's Project A finds little gain in validity for test dimensions already covered by the ASVAB (L. Hanser, personal communication, November 1986), but substantial increases in validity for non-ASVAB content domains, such as temperament, psychomotor, and vocational interests.

Interest in expanding the content domains being tested has been high in the last few years. This interest is being addressed in research efforts by the U.S. military services. The notable programs are the Air Force Project LAMP, the Army Project A, and new predictor development efforts by the Navy.[1]

Project LAMP

The U.S. Air Force Learning Abilities Measurement Program (Project LAMP) is designed to improve selection and classification in the Air Force. Its goals are to denote the basic parameters of learning ability, develop techniques to assess cognitive ability, and investigate the feasibility of applying a cognitive model-based system to psychological assessment (Kyllonen, 1985).

Information Processing Speed. Timed performance on simple tasks is associated with general intelligence (Jensen, 1982). Two approaches are used for studying processing speed: the cognitive correlates approach and the cognitive components approach (Sternberg, 1981; Pellegrino and Glaser, 1979). In the cognitive correlates approach, individuals attempt tasks which are hypothesized to be the basic mental operations (e.g., choice reaction time) underlying more complex cognitive processing. Project LAMP utilizes several types of information processing tasks: choice reaction time, letter matching, and memory scanning (Kyllonen, 1985). These tasks represent a hetrogeneous mix of information processing skills that are related to mental ability (Lansman, Donaldson, Hunt, and Yantis, 1982). In the cognitive components approach (Sternberg, 1981), individuals attempt tasks which are similar to those found in standard intelligence tests (e.g., verbal analogues). Models representing different strategies of test performance are tested by systematically varying characteristics of the test

---

[1] A detailed discussion of these programs can be found in Park (1988).

and observing the effects of the presence or absence of these characteristics on test performance. Many of the Project LAMP tests involve the paired-associates paradigm in which the memory load, and visual or verbal complexities are systematically varied (Christal, 1985; Kyllonen, 1984). Here the object is to isolate the particular information processing components that are fundamental to task performance.

Project A

The U.S. Army is developing an expanded predictor battery under the auspices of Project A (Campbell, 1986). The goal of Project A is to evaluate and update the selection and classification procedures in the Army. Under Project A, a comprehensive set of job performance measures has been developed to assess the validity of the ASVAB along with a set of experimental predictor measures. These tests are a spatial ability test battery, inventories measuring temperament, biodata, and vocational interests, and a battery of computerized perceptual/psychomotor abilities.

Spatial domain. The Project A paper-and-pencil spatial ability battery covers three facets of spatial ability: spatial orientation, or the ability to imagine the appearance of an object from several perspectives; spatial relations, or the ability to mentally move objects, such as mental rotation; and spatial visualization, or the ability to imagine the relative movements of several objects in a complex visual task.

Computerized Perceptual/Psychomotor. Project A incorporated the cognitive correlates approach in the computerized assessment of the following construct domains: perceptual speed and accuracy, memory, and reasoning. These domains use the reaction time paradigm (Sternberg, 1981), which collects time measures as the dependent variable. Reaction time paradigms which are associated with the cognitive correlates approach are: choice reaction time, letter matching, memory search, and sentence verification.

The Project A computerized battery was also developed to measure three psychomotor ability constructs (Imhoff and Levine, 1981): precision/ steadiness, multilimb coordination, and movement judgement. The battery involved such tasks as tracking a moving target or projecting a missile toward a moving target.

Temperament/Biodata. Project A included the measurement of temperament, biodata, and interests as two noncognitive inventories, the Assessment of Background and Life Experiences (ABLE) and the Army Vocational Interest Career Examination (AVOICE). The ABLE measures selected criterion-revelant attributes of temperament, and assessed life history and Physical Condition. The AVOICE is a version of the U.S. Air Force's Vocational Interest Career Examination (VOICE) (Alley and Matthews, 1982), modified to be content valid for Army occupations. It measured general interests, expressed interests, and compatibility with the organizational environment.

Navy Program

Within the context of new test development, the U.S. Navy is assessing a variety of predictors for job performance measurement. These tests are mostly computerized spatial, perceptual, and cognitive processing tests and include a paper-and-pencil biodata inventory. The Navy work program is directed into four distinct efforts: the Graphic Information Processing Tests (GRIP), Spatial Abilities, Cognitive Speed, and the Armed Services Applicant Profile (ASAP) inventory.

GRIP. The computerized GRIP battery was originally developed as task analogues to work samples (Cory, 1977). The focus was on simulating tasks in job performance for the purpose of developing training simulation systems.

Spatial Tests. Under the auspices of the Computerized Testing Technology project, static and animated tests of spatial-visual reasoning have been developed to fully capitalize on the computer's capability for test administration (Wolfe et al., 1987).

Cognitive Speed. The Cognitive Speed project is investigating information processing speed as a distinct dimension correlated to general intelligence (Saccuzzo, Larson, and Rimland, 1986).

ASAP. The Armed Services Applicant Profile (ASAP) is a biodata instrument containing a heterogenous set of empirically keyed items. It was designed as a screen for selection into military enlisted service, to substitute for the selection standard of a high school diploma (Trent, 1987).

Other Service Activities

Predictor development remains a Service-level activity. It is the Services' responsibility to provide the data on any proposed test for the FT Committee to evaluate. The criteria that the committee applies in evaluating the candidate test are shown in the November 1986 minutes in Appendix A, and in Attachment A as the "ACAP Future Tests Check List". In addition, the software for the test should be available to the committee for evaluation.

The Services have developed software for running a number of their own predictors on the ACAP equipment. The largest effort has been carried out by Navy Personnel Research and Development Center (NPRDC), in which 18 predictors and a stand-alone driver have been programmed (J. Wolfe, personal communication, January 1988). The Job Performance Measurement (JPM) Working Group sponsored an effort to convert seven spatial tests and develop a driver (R. Harris, personal communication, April 1988). Finally, the Army Research Institute (ARI) has converted two spatial tests to the system.

Validation efforts on new predictors are a mixed bag. The largest effort has been by ARI, a concurrent validation of the Project A test battery against job performance measures in 19 representative Military Occupational Specialties (MOS) (N = 9,300). Although ARI plans to adapt

Project A tests to the ACAP system, no Army validity studies have been carried out using the system. NPRDC has pilot tested their new predictors on the ACAP system, and data collection is underway to validate a trimmed predictor battery (D. Alderton, personal communication, May 1988). Both NPRDC and the Air Force Human Resources Laboratory are concentrating on validating the JPM test battery on stand-alone HP-IPCs. Data have been collected on the seven JPM spatial tests and Service specific performance and training measures. The data are in the process of being analyzed. The Center for Naval Analyses is undertaking its own JPM study on four Occupational Areas (OA). Data analysis for the Infantry OA is underway (P. Mayberry, personal communication, May 1988).

EXPANDABILITY OF ACAP

The hardware and software of ACAP impose several constraints on the kinds of tests that could supplement the system. Both will be discussed below.

Computerized ACAP testing has two modes of operation. First, the Local CAT-ASVAB Network (LCN) mode of operation simultaneously administers tests to applicants stationed at several Examinee Test (ET) Stations. Second, the stand-alone mode of operation is required in case the LCN fails to operate properly, or when only one examinee is being tested. The functional requirements of the ET stations in the networking and stand-alone modes of operation were thoroughly described earlier by Jones-James and Rafacz (1985) and are summarized in Park (1987) (Attachment B).

The design concept for the LCN mode has the software and item pool being loaded into the Test Administrator's (TA) Station before being broadcast, subtest by subtest, out to the ET Stations. In the standalone mode, the TA Station functions as an ET Station. One hardware constraint is the RAM limits of the system. In the LCN mode, the TA and the ET Stations hold 2.5 and 1.5 megabytes of RAM, respectively. Under the current design concept, the CAT-ASVAB uses nearly the full RAM capacity of the system (B. Rafazc, personal communication, January 1988).

Another hardware constraint is the ACAP examinee input device. This device precludes using tests such as those found in the Project A computerized psychomotor battery. These psychomotor tests require special equipment with the appropriate levers and knobs. Moreover, the equipment requires regular and careful maintenance. Particular attention needs to be paid to the calibration of any moving part of the equipment that the examinee uses.

In addition to constraints imposed by hardware, the compatibility of software is an issue. Any "future test" (FT) must be integrated with the ACAP testing system so as to allow the testing flow to transition smoothly between the CAT-ASVAB and the FT battery. This involves the ACAP software initiating the "future test" by uploading the instruction set and test items, collecting the data, test-interrupt and 'help' function features, and downloading the data.

Several design configurations for adding a FT battery to the ACAP LCN have been discussed in greater detail elsewhere (Park, 1987). These designs either: (1) treat the FT as another ASVAB-like subtest module within the testing system; (2) treat the FT as other ASVAB-like subtest modules, but require the capability to provide a second broadcast of the FTs and overwrite on RAM; and (3) treat the FTs as an independent battery that does not interact with the LCN. The first design may be workable with only one or two tests using the current testing system. It is meant to require minimal software revisions, but the RAM restrictions of the hardware severely limit the selection of tests. In the second design, the FT battery is just as much a part of the ACAP as is the CAT-ASVAB. This design would require additional software support resources, and another 1 megabyte of RAM on the TA Station. The third design has the FT battery reside on separate disks at the ET Stations. This minimizes the software integration support required of NPRDC. In all cases, system testing and analyzing the impact on changing operational testing procedures would need to be done.

DISCUSSION

The efforts of the U.S. military Services to augment the ASVAB have been reviewed. Common among all projects is the concern for attributes beyond verbal ability, quantitative ability, and technical knowledge. The principal approach has been to examine content domains not currently in the ASVAB for their potential to increase validity for predicting training or job performance. These non-ASVAB domains include temperament and biodata content areas; and spatial, perceptual, and psychomotor abilities.

Integration of the FT systems into the ACAP system is a real problem. This problem can be broken into concerns about: (1) software, (2) hardware, and (3) dedicated time.

(1) Software. Systems integration of FTs with the ACAP software is best done by the NPRDC programming group which controls the ACAP software. However, programming the FT battery competes with the other software development demands on the project. Thus, the Services which contribute to the FT battery should consider easing the burden by developing functional specifications from which the FT battery may be programmed, and by providing a form of the test already coded.

(2) Hardware limits. If the current ACAP system has insufficient RAM to include a FT battery, alternatives must be considered. These include: increasing RAM, rewriting the software, or administering the FT battery separately.

(3) Time limits. The maximum time limit for CAT-ASVAB with seeded items should not exceed the paper-and-pencil (PP) ASVAB testing time. Based on data presented at the April 1988 Psychometric Committee meeting, the CAT time is nearly that of the PP time. Thus, little time is available for administering FTs. If ASVAB and FTs are given in the same testing session, either ASVAB subtest times must be shortened, or the total testing session lengthened.

REFERENCES

Alley, W.E. and Matthews, M.D. (1982). The Vocational Interest Career Examination: A description of the instrument and possible applications. The Journal of Psychology, 112, 169-193.

Campbell, J. P. (1986). Validation analysis for new predictors (RS-WP-86-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Christal, R.E. (1985). New cognitive tasks being evaluated by TTCP Services. Unpublished manuscript.

Cory, C.H. (1977). Relative utility of computerized versus paper-and-pencil tests for predicting job performance. Applied Psychological Measurement, 1, 551-564.

Divgi, D.R. (1986). On some issues in the Accelerated CAT-ASVAB Project. CRM 86-231, CNA Research Memorandum, Alexandria, VA.

Imhoff, D.L., and Levine, J.M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHRL TR-80-27). Brooks AFB, TX: Manpower and Personnel Division, U.S. Air Force Human Resources Laboratory.

Jensen, A.R. (1982). Reaction time and psychometric g. In H.J. Eysenck (Ed.), A model for intelligence. Pringer-Verlag.

Jones-James, G., and Rafacz, B.A. (1985). Accelerated CAT-ASVAB Project (ACAP) Development System Functional Requirements, Addendum #2: Examinee Test Station. San Diego, CA: Navy Personnel Research and Development Center.

Kyllonen, P.C. (1985). Theory-based cognitive assessment (AFHRL-TP-85-30). Brooks AFB, TX: Manpower and Personnel Division, U.S. Air Force Human Resources Laboratory.

Kyllonen, P.C. (1984). Dimensions of information processing speed (AFHRL-TP-84-56). Brooks AFB, TX: Manpower and Personnel Division, U.S. Air Force Human Resources Laboratory.

Lansman, M., Donaldson, G., Hunt, E., and Yantis, S. (1982). Ability factors and cognitive processes. Intelligence, 6, 139-160.

Moreno, K.E., Wetzetl, C.D., McBride, J.R., and Weiss, D.J. (1983). Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and Computerized Adaptive Testing (CAT) subtests. (NPRDC-TR-82-27). San Diego, CA: Navy Personnel Research and Development Center.

Park, R.K. (1987). Accelerated CAT-ASVAB Project (ACAP) Development System Functional Requirements, Addendum for Examinee Test Station: Supplemental Test Battery (RS-WP-87-04). Alexandria, VA: U.S. Army Research Institute.

Park, R.K. (1988). Augmenting the ASVAB. Unpublished manuscript.

Peterson, N.G. (ed.)(1987). Development and Field Test of the Trial Battery for Project A (TR 739). Alexandria, VA: U.S. Army Research Institute.

Pellegrino, J.W., and Glaser, R. (1979). Editorial: Cognitive correlates and components in the analysis of individual differences. Intelligence, 3, 187-216.

Saccuzzo, D.P., Larson, G.E., and Rimland, B. (1986). Speed of information processing and individual differences in intelligence (TR 86-33). San Diego: Navy Personnel Research and Development Center.

Sternberg, R.J. (1981). Testing and cognitive psychology. American Psychologist, 36, 1181-1189.

Trent, T. (1987, August). Armed Forces Adaptability Screening: The utility of the biographical inventory. Paper presented at the annual meeting of the American Psychological Association, New York City, NY.

Wolfe, J.H., Alderton, D.L., Cory, C.H., and Larson, G.E. (1987). Reliability and validity of new computerized ability tests. Unpublished manuscript.

APPENDIX A

MINUTES TO THE
CAT-ASVAB FUTURE TESTING COMMITTEE

MINUTES

CAT-ASVAB FUTURE TESTING COMMITTEE

10 February 1986

1. On 10 February 1986, a CAT-ASVAB Future Testing Committee meeting was held at the Navy Personnel Research and Development Center, San Diego, California. Attendees are shown at Attachment 1.

2. The meeting was chaired by Dr Randolph Park, Army Research Institute (ARI). The other Service representatives were: Dr Lonnie Valentine, Air Force Human Resources Laboratory (AFHRL); Dr Paul Mayberry, Center for Naval Analyses (CNA); and Dr Robert Morrison, Navy Personnel Research and Development Center (NPRDC).

3. The following topics were discussed.

a. The ongoing programs and plans of the four Services was reviewed. The briefing slides are at Attachment 2. ARI is pursuing the transfer of Project A tests onto the ACAP System. AFHRL is developing new tests from BAT and the LAMP project for operational use. NPRDC is also developing new tests. CNA is interested in obtaining tests being developed by the other Services for data collection purposes.

b. A general consensus of the committee members was that information on Service-specific development of "new predictors" should be shared with the other Services. The concern here is to coordinate the Services' efforts in an attempt to reduce redundancy.

c. The committee members agreed to strive to provide recommendations on promising predictor tests for the purposes of selection and classification. The point at which the committee would feel ready to make these recommendations and the mechanics of how the committee would arrive at the recommendations will be discussed in future meetings.

d. It was agreed that before plans could be made as to where the committee is going, we need to identify what is going on. That is, have an audit of what is currently available. This involved a discussion of tests, evaluation criteria (redundancy), type of information being collected (standards or statistics, data collected), relation to MOS, and priority standpoint. From this we could make recommendations on the most promising tests. It was agreed that it would be expediate for a Contractor to accomplish the audit. Dr Laabs, NPRDC, is currently obtaining this information through an effort with ARO. The committee made recommendations about possible evaluation criteria that would be of interest, and identified academic individuals who could accomplish the effort. Dr Park was tasked to meet with Dr Laabs to express the committee's recommendations and concerns. The SOW was then made available to the other committee members for review and comments. Some committee members also expressed interest in being available for the kick-off meeting on the Contract. The Contract is expected to be let in April 1986 and the audit information made available to the committee in September 1986.

CAT-ASVAB FUTURE TESTING TECHNOLOGY COMMITTEE

8-9 May 1986

1. On 8-9 May 1986, a CAT-ASVAB Future Testing Technology Committee meeting was held at the Center for Naval Analyses, Alexandria, Virginia. Attendees are shown at Attachment 1.

2. The meeting was chaired by Dr Randolph Park, Army Research Institute (ARI). The other Service representatives were: Dr Lonnie Valentine, Air Force Human Resources Laboratory (AFHRL); Dr Paul Mayberry, Center for Naval Analyses (CNA); and Dr Robert Morrison, Navy Personnel Research and Development Center (NPRDC).

3. The following topics were discussed.

a. The general consensus of the FTT committee denoted the experimental battery as a complementary supplement to the ASVAB to support differential classification. Validity to training criteria will be improved by measuring abilities not tested now by the ASVAB. The nature of the experimental battery will include spatial, perceptual, and psychomotor tests.

b. The committee discussed technical issues and concerns on the implementation of an experimental battery for the Accelerated CAT-ASVAB Project (ACAP). We met to devise a strawman implementation plan for an experimental test battery on the ACAP system. The strawman plan delineates four phases. Phase I emphasizes the investigation of multiple batteries by the specific services. Data that will be examined are: correlation with ASVAB, correlation with training measures, correlation with job perfomance, correlations with evaluation/performance rating, and uniqueness. Phase II commences with ACAP. Here, proposed experimental batteries are rotated as available and when sufficient data has been collected. Phase III commences with FSD. Presumably by now, enough data exists for the committee to reach a general agreement on a large set of tests to be used as a supplement to the CAT-ASVAB. Rotation of parts of the test will continue. Phase IV is the final set to be used to support differential classification by each Service as they choose.

c. The committee recommends that 90 minutes be dedicated to the experimental battery.

d. Dr Mayberry, CNA, briefed an update on the CNA JPM contract. A decision for the winning contractor had just been made.

e. Dr Petersen, PDRI, briefed the conceptual design of the predictor battery for ARI Project A. ARI expects a draft of the TP explicating the development of the trial battery by the end of May 1986, and a draft of the results of the concurrent validation of the predictor battery by the end of Sep 1986. These documents will be distributed to the committee when available.

f. The ARO contract, intended for an audit of new predictor tests being examined by the Services, is looking for an academic principal investigator. Both Bob Linn and Lyle Jones were approached at AERA in San Francisco to be PI for the contract. Additional candidates for the contract were submitted by the committee to Dr Laabs. The committee expressed a general preference for an academician to do the work.

g. Mr Wolfe, NPRDC, presented the Navy's selection of 5 "best" tests to be included in the experimental battery. AFHRL and ARI will strive to have recommendations for their own Service-specific tests by the next FTT meeting.

h. By unanimous vote, "the FTT committee of the CAWG will act in an active advisory capacity and is interested in the accomplishment of the NPRDC SOW entitled 'Investigation and evaluation of new computerized predictor measures.' Recommend the CAWG evaluate the SOW for formal endorsement."

i. By split vote (3-Yes, 1-Abstension), "the FTT committee of the CAWG will act in an active advisory capacity and is interested in the accomplishment of the follow-on concurrent validation work proposed by NPRDC (ltr Ser 62/325, 25 Apr 86, para 4). Recommend the CAWG evaluate the proposal for formal sponsorship."

j. The next meeting of the FTT committee will be held 24-25 July 1986 in San Antonio, TX, immediately following the JPMWG meeting.

MINUTES

CAT-ASVAB FUTURE TESTING TECHNOLOGY COMMITTEE

6-7 August 1986

1. On 6-7 August 1986, a CAT-ASVAB Future Testing Technology Committee meeting was held at the Air Force Human Resources Laboratory, Brooks AFB, Texas. Attendees are shown at Attachment 1.

2. The meeting was chaired by Dr. Randolph Park, U.S. Army Research Institute (ARI). The other service representatives were: Dr. Lonnie Valentine, Air Force Human Resources Laboratory (AFHRL); and Dr. Robert Morrison, Navy Personnel Research and Development Center (NPRDC).

3. The following topics were discussed:

a. The committee discussed technical issues and concerns on the implementation of the experimental battery for the Accelerated CAT-ASVAB Project (ACAP). Phase I involves the investigation of service specific batteries by each of the services. There was general consensus that the collection predictor and training criterion data would be best accomplished at the RTCs during this phase.

b. Dr. Woltz, AFHRL, presented the Air Force's selection of candidate tests for the experimental battery. Dr. Park, ARI presented the Army's selection.

c. The committee discussed the limitations of the current ACAP response device. The response device does not accommodate the Army's tests.

d. The status of the different service efforts related to new predictor development was discussed. The committee agreed to send timetables of projects to be placed under one document by the chairman. This information will include when validation analyses will be accomplished, data collection on HPs and other systems, and conversion of tests to the HP.

3. The ARO contract, intended for an audit of new predictor tests in the Services, has been let to Dr. Earl Hunt, University of Washington. Dr. Hunt is contacting the Services for material to review.

MINUTES OF THE 24-25 NOVEMBER 1986
CAT-ASVAB FUTURE TESTING COMMITTEE MEETING

1. The meeting of the CAT-ASVAB Future Testing Committee was held 24-25 November 1986 at the Center for Naval Analyses in Alexandria, VA. The agenda is shown at Attachment 1, and a list of attendees is shown at Attachment 2. A discussion of topics precedes presented materials at Attachment 3.

All FT Committee recommendations made below were unanimous.

2. Preparation of Test Material/SW
The FT Comittee recommends that the individual Services be responsible for programming their own FT as separate modules that can be readily integrated into a driver program. It is desirable that the subtest modules follow a standardized I/O format. The minimum data needed for a FT subtest are: subtest code ID, item code ID, response latency, and examinee response option. The subtest SW shall be transportable to the other Services. Source code shall be provided. Documentation shall be provided and will minimally include a list of subprograms, the functions of subprograms, and the source code in a modular programming format.

3. Collection of Predictor/Criteria data at the Service level
The FT Committee strongly recommends that FT tests and the CAT-ASVAB be given concurrently. The CAT-ASVAB should consist of the operational SW and item pool. The experimental item pool is not adequate because of the change in the nature of the speeded subtests and the lack of an equating for the power subtests. Only one form of the CAT-ASVAB item pool is needed along with the HP standalone SW. The block seeding of items should be preferably switched off.

4. Other SW support to FT
The FT Committee recommends that the CAWG approved ACAP software for reading the Master Datafile by CAMP be made available for use by the FT Committee Service representatives.

5. Criterion to evaluate Future Tests
The FT Committee recommends the following criteria be used to evaluate candidate FT material:

a) uniqueness (test-retest Reliability) - (mr-squared)
b) Incremental validity (w/ training and w/ JPM)
c) Reliability
d) Coachability
e) Practice effects (from retest data)
f) Apparent testing strategies
g) Concurrent validity w/ ASVAB
h) Age of ASVAB scores when concurrent validity not available
i) SW availability
j) Time length (95% completion rate)
k) RAM requirements
l) Cost/benefit tradeoff of j) and b)

m) Subgroup differences
n) Differential validity by MOS
o) Ease of conversion to production tests (FSD consideration only)
p) Fits with standardized response device (not specified for FSD)

In addition to the above criterion, filler tests
r) Should not distract from dedicated tests (not interesting/not boring)
s) Be easy to extend (variable test length)
t) More than one variable length filler test

6. Privacy Act
It is noted that for the FT battery, the examinee should be informed that their participation is voluntary and they may have the option of sitting without disturbing people. The issue here is when the Privacy Act should be given.

7. Subtest Format
The FT Committee recommends that for the purposes of ACAP, all FT subtest modules be in HP-UNIX "C". An exception is that other languages are acceptable as long as no additional filehandling problems are created.

8. Driver Program for FT
The FT Committee recommends that the ACAP SW must initiate the driver program for FT.

9. Driver Program Format
It is noted that the following features of the driver program should be kept in mind:
a) general administrative instructions, including Privacy Act
b) define subtest order
c) check system time
d) choose filler order if time is short
e) bridge subtests
f) automate test identification
g) open/close datafile
h) stop timer and negate item when 'help' function is depressed
i) close out testing
j) send FT data to Master Datafile for examinee behind ASVAB

10. Information request for Code 63
The FT Committee needs to know the amount of RAM available for FT use. This will determine whether an intelligent driver or chaining is used. The committee also needs to know: the I/O conventions, ease of replacing seeded blocks with FT, placement of FT on the backup disk.

11. Dedicated Examinees for FT
The FT Committee recommends that at least one-third of the SE and IOT&E samples be reserved for FT rather than item seeding.

12. Dedicated Time for FT

The FT Committee recommends that at least 35 minutes be reserved for FT.

13. Administration of Subtests

The FT Committee discussed the subtest administration procedure. At the end of the CAT-ASVAB, there should be an autoexec routine to a) clear RAM, b) read and display the instruction set, and c) load the FT tests. As the instructions are displayed, the examinee reads the instructions as the tests are loaded into RAM. FT subtests can be loaded from one of three sources: a) the TA Station, b) the backup disk, or c) another disk. The loading sources are discussed in order of preference by the FT Committee.

Option A: FT subtests are loaded form the TA Station. This will require complete access to the CAT-ASVAB SW and would be most efficiently be accomplished by NPRDC (Code 63). By completely integrating the FT into the CAT-ASVAB SW and CAT systems SW, little additional demand is made on the TA.

Option B: FT subtests are loaded from the backup disk. This will require systems integration support from NPRDC (Code 63). Little additional demand is made on the TA.

Option C: FT subtests are loaded from another disk. This option will require less support than the above options from NPRDC (Code 63). The CAT-ASVAB SW will need to tell the examinee what to do after the ASVAB is completed. The TA will be required to change disks. Possible problems arise if: more then one person finished or needs help at the same time, losing disks, and merging data if merging is not accomplished at the TA Station.

14. Downloading data

It is noted that the data from dedicated FT subtests should be saved and downloaded for transmission to CAMP. Three downloading options are forwarded and discussed in order of preference by the FT Committee.

Option A: The data is downloaded at the tailend of regular CAT-ASVAB testing. This will require systems integration support from NPRDC (Code 63) at NPRDC.

Option B: The data should be match/merged at the TA Station. This will require systems integration support from NPRDC (Code 63).

Option C: The data will be collected on separate disks. Minimal SW support will be required by NPRDC (Code 63). A separate driver program on an independent disk can handle this. The procedure will require match/merging of FT subtests to the ASVAB later, probably at CAMP.

15. Multiple Forms

It is noted that multiple forms of dedicated FT subtests are created by changing the order of FT to be presented.

16. FT Committee Response to CAWG Tasking

The FT Committee chairman tasks each of the Service representatives to submit the below information on all tests that are reasonably expected to be programmed over to the HP by the ACAP Score Equating phase. Each Service representative should forward his submission to the FT Committee chairman NLT 1 Jan 87.

a) Uniqueness
b) Incremental validity
c) Reliability
d) Sample sizes
e) Mean age of ASVAB scores
f) Actually programmed/expected date of completion
g) Time length of testing
h) Time length of instructions
i) Number of items
j) Description of tests (for coachability)
k) Subgroup differences w/ minority sample size
l) MOS differential validity should be available

17. Current Plans of Action

NPRDC (Code 622) has a contract to convert 20 Navy FT over to the HP. Accomplishment of this task is expected NLT 30 Jun 87. A program driver will also be accomplished under the contract if time allows.

ARI (PERI-RS) has a contract to convert 3 Army FT over to the HP. Accomplishment of this task is expected NLT 30 Apr 87. ARI (PERI-RS) has an inhouse effort to convert 2 Army FT and a driver program over to the HP. Accomplishment of this task is expected NLT 30 Apr 87.

18. The FT Committee tasks the FT chairman to submit a memorandum in response to a Sep 86 inquiry by Dr Lancaster (OASD) on HW support for FT validity data collection.

Minutes of the 7 July 1987 Meeting
of the
Future Testing Committee

1. The CAT-ASVAB Future Testing committee met on 7 July 1987 in San Diego, CA. A list of attendees is shown at Atch 1. The agenda is shown at Atch 2.

2. Service research and procurement efforts were reviewed. ARI is currently procuring 25 HP Int PCs. Five will be used for development, and 20 will be used for testing. Delivery is expected during the Aug-Sep timeframe. Project A Spatial Reasoning and MAP tests have been converted to the HP, and will be available on request. ARI anticipates award of a larger contract to convert Project A and other Future Tests (FTs) by Spring 1988. NPRDC currently owns 3 HPs for Future Testing programming. A procurement of 30 HPs (8 development/22 testing) is expected for delivery in Aug 1987. The delivery of 2 MG RAM boards is expected in Oct 1987. Upwards of 20 FTs are now programmed on the HP and are being tested at the RTC. A GSA contract will be started for integrating the 20 tests. CNA is willing to use any promising tests available, but has no current plans for FT development or testing. AFHRL has nor current plans for FT conversion to the HP.

3. The candidate FT list shown at Atch 3 was reviewed. The interactive screen dialoges (ISDs) for most of the tests were collected. These will be reviewed by the chairman and returned to the sponsor Service for revision NLT 7 Aug 1987.

4. The FT designs from the ARI working paper (WP-87-06) were reviewed. The consensus of the committee was that Design 1 should be given further consideration. Design 1 treats the FT as added subtest modules. When the CAT-ASVAB is initially broadcast at the beginning of the CAT test administration, the FT are also broadcast to the ET Stations. The CAT-ASVAB is administered, followed by the FT. When test administration is completed, the CAT-ASVAB and FT are downloaded to the TA Station. The FT are treated as part of the examinee database.

This design is limited to the amount of RAM residual after the CAT-ASVAB and on the amount of time to be dedicated to FT. The committee discussed the use of 2 MB RAM boards inserted into the ET stations, but have questions about the dollar cost and reliability.

5. Each of the Services are responsible for providing the (1) data requirements and addressing the elements in the (2) FT Checklist for each test submitted to the FT list. These items will be provided to the chairman NLT 20 Aug 1987.

Minutes of the 19 January 1988 Meeting
of the
Future Testing Committee

1. The CAT-ASVAB Future Testing committee met on 19 January 1988 in San Diego, CA. A list of attendees is shown at Atch 1. The agenda is shown at Atch 2.

2. Service research and procurement efforts were reviewed. ARI has procured 25 HP Int PCs. ARI anticipates award of a larger contract later this year to convert Project A and other Future Tests (FTs). NPRDC owns 30 HPs for Future Testing. About 20 HPs lack RAM boards. A GSA contract was accomplished to integrate the NPRDC's future tests.

3. AFHRL, in a cooperative study between the MO and ID divisions, has collected experimental future test data on 142 Air Force personnel. The test measures involved are the 7 tests programmed in "C" under Dick Harris' contract with Eagle Technology. The subjects are 69 Airmen working in the 732X0 speciality and 73 Airmen working in the 423X5 speciality. All 142 cases had been subjects for WIPT assessment in ID's work in performance measurement in their respective specialities. AFHRL is assembling a datafile on these cases which will include: (a) identification data; (b) experimental test item responses and scores; (c) ASVAB scores; (d) job performance measures. AFHRL will analyze these data to ascertain (a) ASVAB validity for job performance, (b) experimental test validity for job performance, (c) experimental test unique validities in the context of ASVAB. AFHRL reports two observations from their data collection experience: (a) transportation of equipment to field sites is cumbersome. AFHRL carried equipment as "carry-on" baggage. New airline regulations will be permit this in the future, so equipment will have to be shipped in advance. (b) problems with equipment reliability, and time is required for setup and checkout at new locations.

3. The candidate FT list shown at Atch 2 was reviewed. The interactive screen dialoges (ISDs) for these tests were forwarded to Bernie Rafacz.

4. The Services are asked to provide summaries of updated technical information for each candidate test. The summaries from Jan 87 were passed out as guidance.

5. A draft of a briefing on future testing was reviewed. Many suggestions were made by the committee and will be implemented by the chair.

ATTACHMENT A

ACAP FUTURE TESTS CHECK LIST

ACAP FUTURE TESTS CHECK LIST, April 1987

Test Administration

A. Display format
   1. similarity to CAT ASVAB:
   2. character size: like the CAT ASVAB
   3. font: like the CAT ASVAB
   4. display of time remaining in subtest: bottom - toggle? must be a time limit.
   5. display of number of items remaining: bottom
   6. number of items per screen: 6b. number of screens per item
   7. split screen
   8. area and location of graphics
   9. moving graphics

B. Speed of display
   1. maximum delay for item presentation:
   2. variability of delay between items: low
   3. variability of delay between subtests:
   4. variability of delay between CAT ASVAB and FT battery
   5. time to proctor call: same as in pretest, rounded up to nearest 10 seconds.

C. Examinee input device
   1. keyboard overlay
   2. response confirmation:

D. Rest periods
   1. during subtests:
   2. between subtests:

E. Omitted responses
   1. how to score

F. Order of subtests
   1. ASVAB before blocks/FT
   2. dedicated tests before filler
   3. multiple forms

G. Maximum time limits
   1. Dedicated time: 35 minutes
   2. Subtest time limit
   3. Time per item
   4. CAT or S controlled screen advancing
   5. out-of-time stopping option

H. Resumption of interrupted subtest
   1. return to interrupted screen
   2. data used/not used in scoring
   3 [illegible]

I. Procedure at end of subtest testing
   1. check time for more dedicated tests
   2. check time for filler tests

ACAP FUTURE TESTS CHECK LIST (cont.)

Implementation

A. Initial operational test and evaluation (IOT&E)
   1. subject requirements
   2. test administration requirements
   3. data requirements

B. Checking performance of CAT system
   1. FT impact on system performance
      -- keyboard
      -- TA station
      -- ET station
      -- interface with System 80
      -- production of final scoring report

C. Requirements for post-IOT&E operational use
   1. Adding new subtests to battery
   2. checking validity of FT
      -- corr/ with job performance
      -- corr/ with training
      -- corr/ with ASVAB
      -- corr/ with other FT
   3. checking reliability of FT
      -- test-retest
         retest interval
         sample
      -- reliability estimate for subgroups and full-range population
   4. checking inclusion of FT with service composites

Item pools

A. Number of items

B. Criteria for item inclusion
   1. inspection of subtest format by service technical reps
   2. selection by service sponsor of FT
      -- inspection of item content
      -- pt-bis corr/ of response alternative with total score

C. Item selection
   1. ordered or random within subtest

Generic Sub design

ATTACHMENT B

WORKING PAPER RS-WP-87-04

ACCELERATED CAT-ASVAB PROJECT (ACAP) DEVELOPMENT SYSTEM FUNCTIONAL REQUIREMENTS, ADDENDUM FOR EXAMINEE TESTING STATION: SUPPLEMENTAL TEST BATTERY

Working Paper

RS-WP-87-04

ACCELERATED CAT-ASVAB PROJECT (ACAP) DEVELOPMENT SYSTEM FUNCTIONAL REQUIREMENTS, ADDENDUM FOR EXAMINEE TESTING STATION: SUPPLEMENTAL TEST BATTERY

Randolph K. Park
Selection and Classification Technical Area

May 1987

REVIEWED BY:
DEIRDRE KNAPP
Selection and Classification
Technical Area

APPROVED BY:
LAWRENCE M. HANSER
Chief, Selection and
Classification Technical Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

TABLE OF CONTENTS

LIST OF TABLES

ACCELERATED CAT-ASVAB PROJECT (ACAP) DEVELOPMENT SYSTEM
FUNCTIONAL REQUIREMENTS
ADDENDUM FOR EXAMINEE TESTING STATION:
SUPPLEMENTAL TEST BATTERY

Introduction

The mission of the CAT-ASVAB Project is to field a nationwide testing system to administer the ASVAB as a computerized adaptive test (CAT). In 1983, this project was directed by the Department of Defense to become the Accelerated CAT-ASVAB Project (ACAP). The ACAP is a limited scale "lessons-learned" project, required to use off-the-shelf equipment to meet an accelerated implementation date. The ACAP is required to administer the computerized tests to civilian applicants for military enlistment for the purposes of selection and classification. Field testing will take place at selected Military Entrance Processing Stations (MEPS) and the associated Mobile Examining Team (MET) sites.

Computerized testing will have two modes of operation. First, the Local CAT-ASVAB Network (LCN) mode of operation will simultaneously administer tests to applicants stationed at several Examinee Test (ET) Stations. Second, the stand-alone mode of operation is required in case the LCN fails to operate properly, or when only one examinee is being tested. The functional requirements of the ET stations in the networking and stand-alone modes of operation were thoroughly described earlier by Jones-James and Rafacz (1985) and will be summarized below.

This document is to be included in the series of supplementary documents intended to amplify the ACAP, Functional Requirements Document (Rafacz and Tiggle, 1985). This addendum will describe the LCN and stand-alone ET Station functional requirements for the administration of a post-CAT-ASVAB supplemental test battery. The supplemental test battery will consist of subtests using blocks of ASVAB or non-ASVAB items.

Examinee Test Station Functional Requirements

ET Station LCN Mode

The networking mode of operation consists of several ET Stations linked with a Test Administrator's (TA's) Station. The function of the TA Station is to: (a) download the testing software and data to the ET Stations; (b) monitor all ET Station activity; and (c) upload examinee data and test results. A detailed description of the functional requirements for the TA Station and LCN interfacing is given in Abrams and Rafacz (1985).

The ET Station functions are divided into two major drivers, the Network Boot-up driver and the Network ET Station driver. The generic ET station software requirements are shown at Table 1. The associated software functions are shown at Table 2.

Table 1. Generic ET Station LCN Software Requirements

(Network Boot-up Driver)
- A. Hardware startup and self-test
- B. Data acceptance
    1. ET Station software
    2. Test item banks
    3. Supporting data files

(Network ET Station Driver)
- C. Examinee orientation
    1. Examinee log-on
    2. CAT-ASVAB introduction
    3. Privacy Act Statement
    4. Examinee hardware familiarization
- D. Test administration
    1. General Directions
    2. Test familiarization and practice
    3. Testing process
        - a. item selection and display
        - b. response verification
        - c. scoring a valid response
        - d. estimating ability
        - e. test termination rule
        - f. record response data on backup media
- E. Administering a questionnaire
- F. Personal data receipt from TA Station
- G. Personal data verification
- H. Test data transmission to TA Station
- I. Examinee salutation

Table 2. Generic ET Station Functions

Software functions	Cross-Reference in Table 1
I. NETWORK BOOT-UP program	
A. HARDWARE START-UP	A.
B. ACCEPT ET DATA	B.
II. ET STATION DRIVER program	
A. ACCEPT EXAMINEE IDENTIFYING DATA	
B. EXAMINEE ORIENTATION	C.
C. ADMINISTER TEST	D.
D. ACCEPT PERSONAL DATA	F.
E. VERIFY PERSONAL DATA	G.
F. ARCHIVE DATA	H.

Network Boot-Up Driver Program

Purpose. The NETWORK BOOT-UP program is to perform a self-diagnostic verification on the Hewlett-Packard Integral Personal Computer (HP-Int-PC) via the HARDWARE STARTUP function, and load the networking software and support data into the ET Station RAM area via the ACCEPT ET DATA function.

Process. The TA inserts the NETWORK BOOT-UP driver into the ET Station before it is powered up. When the ET Station is switched on, the HARDWARE START-UP function is automated. When the HARDWARE START-UP function is completed, the TA is queried for the desired operating configuration (Network or Stand-alone). If networking is chosen, the HP-IL networking software is loaded into RAM and activated. (For stand-alone mode, the networking software is skipped. A further discussion of the stand-alone mode is found below) If the LCN link to the TA Station is active, the HP-IL networking software is activated and communications are established. The NETWORK BOOT-UP program is ready to accept data from the TA Station (test administration software, info-tables, item pools). The NETWORK BOOT-UP driver proceeds by activating the ACCEPT ET DATA function. The ACCEPT ET DATA function is to accept the ET Station test administration software and supporting data from the TA Station via the HP-IL interface. If the transmission of the supporting data from the TA station to the ET station is successful, control is passed to the ET STATION DRIVER and the NETWORK BOOT-UP driver terminates. If the transmission is not successful, the TA removes the ET Stations from the LCN. The following software and supporting data now resides at the ET Station and are shown at Table 3.

Table 3. ET Station Resident Software and Data

A. ET STATION DRIVER program
B. HP-IL Networking software
C. Examiner Instruction dialogue
D. Examinee dialogue:
   1. CAT-ASVAB introduction
   2. Privacy Act
   3. Hardware familiarization and practice
   4. Subtest(s) familiarization and practice
   5. Salutation
   6. Error messages
E. Test Identifying data
   1. Form number
   2. Version number
   3. Other
F. Test Data
   1. Info-tables
   2. Operational Item Pools
   3. Experimental Item Pools
   4. Questionnaires

Network ET Station Driver Program

Purpose. The NETWORK ET STATION DRIVER program is to perform CAT-ASVAB test administration.

Input. (a) Date, (b) Station ID Number, (c) Site Location, (d) HP-IL Status Indicator (network or stand-alone), (e) (Test) Form Identifying Data, (f) Examinee Identifying Data (received from TA Station and consists of SSN and biodata), (g) Examinee Response File (uploaded only for RESTART examinees).

Process. The ET STATION DRIVER obtains the above input data from either internal sources or queries the TA. Then the ACCEPT EXAMINEE IDENTIFYING DATA function is called to receive the examinee's identifying data from the TA Station and to create the examinee response file. When the examinee identifying data is accepted, the ET STATION DRIVER examines the examinee response file residing on the ET Disk, to determine if the examinee is NEW or a RESTART.

The program either proceeds to begin NEW examinee introduction procedures, or to RESTART examinee at the beginning of the first noncompleted CAT-ASVAB test. If the examinee is NEW, the examinee response data file is created and holds the date, examinee's identifying data, test identifying data, and reserved area for storing item response data. After successfully creating the examinee response file, the EXAMINEE ORIENTATION function is activated to provide the examinee an introduction to the CAT-ASVAB testing program, Privacy Act information, and the hardware familiarization practice session. Afterwards, the ADMINISTER TEST function is activated.

A RESTART examinee is reassigned to an ET Station by the TA from the TA Station. The RESTART examinee response data file will only be found on the ET Disk. When the examinee is reassigned to an ET Station, the RESTART examinee testing information data is uploaded from the ET Disk into Random Access Memory (RAM). The software testing parameters are initialized to begin testing at the correct position in the test administration process.

For both NEW and RESTART examinees, the ADMINISTER TEST function is activated to administer the CAT-ASVAB testing procedure. The ET STATION DRIVER duplicates the examinee response data in RAM, and updates the ET Disk at the end of each CAT-ASVAB subtest. When the ADMINISTER TEST function is successfully completed, the ACCEPT PERSONAL DATA function will download the examinee's personal data from the TA Station to the ET Station. When personal data is successfully downloaded, the VERIFY PERSONAL DATA function will instruct the examinee to verify the data. When the data have been verified, the examinee will be given a salutation and CAT-ASVAB testing ends.

The ET STATION DRIVER will activate the ARCHIVE DATA function, which merges the examinee personal data and test administration response data files, and then activates the UPLOAD EXAMINEE DATA function, to upload all merged examinee data stored in RAM over to the TA Station. Afterwards, all examinee data are cleared from RAM. The examinee identifier and test session times are appended to the ET Station test session log. The ACCEPT EXAMINEE IDENTIFYING DATA function is called and the ET Station awaits for the next examinee's identifying data to be downloaded.

ET Station Stand-alone Mode

The stand-alone mode of operation consists of the TA Station software, and ET Stations as independent testing units. Both TA and ET functions can take place on the same station, if there is only one station. The TA is required to download examinee data, ET Station software, item banks, and supporting data to each ET Station. At the conclusion of testing, the test data must be manually uploaded to the TA Station.

The ET Station functions are divided into two major drivers, the Stand-alone Boot-up driver and the Stand-alone ET Station driver. The generic ET station software requirements are shown in Table 4.

Table 4. Generic ET Station Stand-alone Software Requirements

---

(Stand-alone Boot-up Driver)
- A. Hardware startup and self-test
- B. Data acceptance
  1. ET Station software
  2. Test item banks
  3. Supporting data files

(Stand-alone ET Station Driver)
- C. Examinee orientation
  1. Examinee log-on
  2. CAT-ASVAB introduction
  3. Privacy Act Statement
  4. Examinee hardware familiarization
- D. Test administration
  1. General Directions
  2. Test familiarization and practice
  3. Testing process
     - a. item selection and display
     - b. response verification
     - c. scoring a valid response
     - d. estimating ability
     - e. test termination rule
     - f. record response data on backup media
- E. Examinee salutation

Stand-alone Boot-Up Driver Program

Purpose. The STAND-ALONE BOOT-UP program is to perform a self-diagnostic verification on the HP-Int-PC via the HARDWARE STARTUP function, and loads the STAND-ALONE ET STATION DRIVER, item banks, and support data into the ET Station RAM area via the ACCEPT ET DATA function.

The main differences between the networking and stand-alone modes for the BOOT-UP driver program are: (a) the TA must manually insert disks to load software, item banks, and supporting data files to require more interactive dialogue; (b) the stand-alone software will perform more verifications for the manual procedures; and (c) the TA must enter a

passcode into each ET Station before the data files can be decoded and loaded into RAM.

Process. The TA inserts the ET BOOT-UP driver into the ET Station before it is powered up. When the ET Station is switched on, the HARDWARE START-UP function is automated. When the HARDWARE START-UP function is completed, the TA is queried for the station type (TA or ET), site number, and the desired operating configuration (Network or Stand-alone). After the information is entered and verified by the TA, the ACCEPT DATA function is activated and queries the TA for the passcode. The ACCEPT DATA function will load the STAND-ALONE ET STATION DRIVER from the ET Disk. The TA is then instructed to remove the ET Disk and insert the appropriate TA Station System Disk containing the item banks and supporting data files. When the data are successfully transmitted, the STAND-ALONE ET STATION DRIVER is activated, and the STAND-ALONE BOOT-UP DRIVER terminates.

Stand-alone ET Station Driver Program

Purpose. The STAND-ALONE ET STATION DRIVER program is to perform CAT-ASVAB test administration.

Input. (a) Date, (b) Station ID Number, (c) Site Location, (d) (Test) Form Identifying Data, (e) Examinee Identifyining Data (SSN), (f) Examinee Response File (uploaded only for RESTART examinees).

Process. The ET STATION DRIVER obtains the above input data from either internal sources or queries the TA. The data is then logged on the ET Disk. RESTART examinees are identified by matching the examinee's SSN to an ID File on the ET Disk. The TA is queried if the RESTART examinee is ready to complete testing. Otherwise, the RESTART examinee can be deleted from the ID File or tested later.

Depending on whether the examinee is NEW or a RESTART, the program either proceeds to give NEW examinee introduction procedures, or to initialize a RESTART examinee at the beginning of the first noncompleted CAT-ASVAB test. For NEW examinees, the TA verifies the examinee's SSN and biodata via menu driven prompts. The examinee SSN, Form number, and NEW=0, entries are created in the ID File. An examinee response data file is created. After successfully creating the examinee response file, the (2) EXAMINEE ORIENTATION function is activated to provide the examinee an introduction to the CAT-ASVAB testing program, Privacy Act information, and the hardware familiarization sequence. Afterwards, the (3) ADMINISTER TEST function is activated.

A RESTART examinee is reassigned to an ET Station by the TA. The RESTART examinee response data file will only be found on the ET Disk. When the examinee is reassigned to an ET Station, the examinee testing information data is uploaded from the ET Disk into RAM. The software testing parameters are initialized to begin testing at the correct position in the test administration process. The (3) ADMINISTER TEST function is activated to administer the CAT-ASVAB testing procedure.

During the ADMINISTER TEST function, the examinee response information is duplicated in RAM and updated on the ET Disk at the end of each CAT-

ASVAB test. When the ADMINISTER TEST function is successfully completed, the examinee will be given a salutation and CAT-ASVAB testing ends.

The ET STATION DRIVER will backup the examinee's data file to the ET Disk. The examinee test session times are appended to the ET Station test session log. The TA backs up the examinee data on the TA Disk, after which the RAM and ET Disks are cleared. The TA is then prompted for another examinee.

ACAP Supplemental Test Battery Functional Requirements

This section will expand the discussion of the ET Station functional requirements to include different design considerations for adding a supplemental test battery to the ACAP. Designs 1, 2, and 3 illustrate different scenerios for the LCN mode of operation. In Design 1, a small number of supplemental tests augment the CAT-ASVAB software and are treated as "extra" subtest modules. In Design 2, the supplemental tests are fully incorporated into the existing network. In Design 3, the supplemental tests are independent of the testing system. Design 4 illustrates a scenario for the stand-alone mode of operation, where the supplemental tests are on an additional disk.

Design 1.

Purpose. The goal of Design 1 is to supplement the CAT-ASVAB with a small number of tests without major reconfiguration of the networking software. The main difference from the networking software that contains only the CAT-ASVAB is that this software will treat the tests from the supplemental battery as added subtest modules.

Process. The TA proceeds to boot-up the testing system as before. The NETWORK BOOTUP DRIVER will proceed with hardware start-up and self-test, and the data acceptance function. During the broadcast of CAT-ASVAB data from the TA Station to the ET Station, the supplemental battery test data will also be transmitted. In addition to the software and support data shown in Table 3, the ET Station will hold the test material associated with the supplemental battery as shown at Table 5.

Table 5. Supplemental Battery Test Material

G. Examiner Instruction Dialogue
H. Examinee dialogue:
    1. Supplemental battery introduction
    2. Privacy Act
    3. Subtest(s) familiarization and practice
    4. Salutation
    5. Error messages
I. Test Identifying data
    1. Test number (Same as f.3, Table 3)
J. Experimental Item Pools (same as f.3, Table 3)

Test administration of the CAT-ASVAB proceeds as previously described. When the CAT-ASVAB has been administered, the ADMINISTER TEST function proceeds with the supplemental battery. The ADMINISTER TEST function will include:

D. Test administration
   1. Supplemental battery introduction
   2. Test familiarization and practice
   3. Testing process
      a. item selection and display
      b. response verification
      c. scoring a valid response
      d. test termination rule
      e. record response data on backup media

A general (D.1, reference above) supplemental battery introduction will be given, followed by the Privacy Act. Like the CAT-ASVAB, each supplemental test will include potentially unique (D.2) test familiarization and practice sequences. Since the supplemental tests are nonadaptive, (D.3.a) item selection is sequential. During the testing process, valid responses are (D.3.b) verified and (D.3.c) scored. The subtest terminates (D.3.d) after the last item is administered, or when the subtest time or total test time expires. Response data are (D.3.e) recorded on backup media at the end of each subtest.

When the ADMINISTER TEST function is successfully completed, the examinee's data will consist of the CAT-ASVAB and the supplemental battery data. The ET STATION DRIVER then proceeds as before with the ACCEPT PERSONAL DATA, VERIFY PERSONAL DATA, ARCHIVE DATA, and UPLOAD EXAMINEE DATA functions.

Design 2.

Purpose. The purpose of this design is to supplement the CAT-ASVAB with a large number of tests that make full use of reconfigured networking software. The main difference from Design 1 is that broadcasting all tests during the ACCEPT ET DATA function will not be possible if the RAM area available on the ET Station is exceeded. The design concept here is that at some point during test administration of the CAT-ASVAB, the RAM area for some of the completed subtests' item pools will be cleared to make room for a later transmission of the supplemental battery. The following scenerio is given as an example of this design concept.

Process. The TA proceeds to boot-up the testing system as before. The NETWORK BOOTUP DRIVER will proceed with hardware start-up and self-test, and the data acceptance function. The CAT-ASVAB and full supplemental test battery are loaded into the TA Station. During the ACCEPT ET DATA function, the ET Station support software will include a DELETE TEST subfunction. During the broadcast of CAT-ASVAB data from the TA Station to the ET Station, one of the supplemental battery tests will be transmitted. The remaining supplemental tests will not be broadcast at this time. Test administration of the CAT-ASVAB proceeds as previously described.

After CAT-ASVAB testing is completed, the DELETE TEST subfunction is activated. The RAM area holding the Mechnical Comprehension subtest is cleared. A transmission is sent to the TA Station to notify the TA that

CAT-ASVAB testing is completed, and the NETWORK BOOT-UP program is transmitted to the ET Station. The NETWORK BOOT-UP driver proceeds to skip the HARDWARE STARTUP function on condition that CAT-ASVAB testing is completed. The ACCEPT ET DATA function is activated. If CAT-ASVAB is completed and supplemental testing is in progress (on the one supplemental test broadcast with the CAT-ASVAB), the ACCEPT ET DATA function is to accept the remaining supplemental battery test administation software and supporting data from the TA Station via the HP-IL interface. If transmission of the supporting data from the TA Station to the ET Station is successful, the NETWORK BOOT-UP driver terminates. If transmission is interrupted by disrupted testing (e.g., the examinee requires a RESTART), the transmission of the supplemental battery is attempted again.

After the one supplemental test is completed, the ADMINISTER TEST function continues with the remaining supplemental tests. As with the CAT-ASVAB subtests, the ET STATION DRIVER duplicates the examinee response data in RAM, and updates the ET Disk at the end of each test.

When the session testing time limit is reached, the ADMINISTER TEST function terminates testing on the current test. The ET STATION DRIVER proceeds as before with the ACCEPT PERSONAL DATA, VERIFY PERSONAL DATA, ARCHIVE DATA, and UPLOAD EXAMINEE DATA functions.

The DELETE TEST subfunction is activated, and the RAM area holding the remaining portion of the supplemental battery is cleared. A transmission is sent to the TA Station to notify the TA that testing is completed, and the NETWORK BOOT-UP program is transmitted to the ET Station. The NETWORK BOOT-UP driver proceeds to skip the HARDWARE STARTUP function on condition that CAT-ASVAB testing is completed. The ACCEPT ET DATA function is activated. If the testing session is completed, the ACCEPT ET DATA function is to accept the Mechnical Comprehension subtest administration software and supporting data from the TA Station via the HP-IL interface. If transmission from the TA Station to the ET Station is successful, the NETWORK BOOT-UP driver terminates. The ACCEPT EXAMINEE IDENTIFYING DATA function is called and the ET Station waits for the next examinee's identifying data to be downloaded.

Design 3.

Purpose. The goal of this design is to supplement the CAT-ASVAB with a large number of tests resident on an independent disk in the LCN mode. Here the design concept is to build an independent supplemental battery that does not require interaction with the network and thereby has minimum software support requirements from the ACAP software programming resources. This design is also applicable for supplementing the CAT-ASVAB in the stand-alone mode.

Process. The TA proceeds to boot-up the testing system as before. The NETWORK BOOTUP DRIVER will proceed with hardware start-up and self-test, and the data acceptance function. Only the CAT-ASVAB software and support data are broadcast from the TA Station to the ET Stations. Test administration proceeds as before.

When the UPLOAD EXAMINEE DATA function is activated, the TA is notified that the examinee personal data and test administration response data files are being uploaded to the TA Station from the ET Disk. When uploading is

completed, the TA manually replaces the ET Disk with the Supplemental Battery (SB) Disk.

Presumably, the SB Disk clears the CAT-ASVAB from the ET Station RAM area. (Note that in this scenerio, subsequent CAT-ASVAB testing is not possible after the first CAT-ASVAB session.) The Input Data (which includes Examinee Identifying Data input when the NETWORK ET STATION DRIVER was initiated), is written onto the SB Disk. The SB Disk downloads the supplemental battery onto the ET Station RAM area. The downloaded software and support data are verified. The supplemental battery will be administered as earlier described. The response data will be recorded on the SB Disk. When the test session time limit is reached, testing will terminate on the current test.

Design 4.

Purpose. Supplement the CAT-ASVAB with a large number of tests resident on a separate disk in the stand-alone mode.

Process. The TA proceeds to boot-up the ET Station in the stand-alone mode, and administer the CAT-ASVAB as previously described. When the examinee completes CAT-ASVAB testing, the TA backs up the examinee data on the TA disk and the RAM and ET Disks are cleared. The TA is then prompted for another examinee or for the supplemental battery. If the TA chooses to test another examinee, CAT-ASVAB testing proceeds as before. If the TA chooses to give the current examinee the supplemental battery, the TA manually replaces the current disk (TA or ET) with the SB Disk. The SB Disk downloads one test at a time onto the ET Station RAM area. The downloaded software and support data are verified. The supplemental test will be administered as described earlier. The response data will be recorded on to the SB Disk. If the test is completed, the test RAM area is cleared and the next test is downloaded. If the test session time limit is reached, testing will terminate on the current test. The TA backs up the SB data on the TA disk, after which the RAM and SB Disks are cleared. The TA replaces the SB Disk with the ET Disk. The TA is then prompted for another examinee.

Discussion

In order for ACAP to meet its mission, several of the Joint-Service approved criteria for evaluating a CAT-ASVAB testing system were dropped. Among these was the requirement for the "expansion-flexibility" of the testing system to accommodate all tests being developed by the Services. By selecting off-the-shelf test administration equipment, the ACAP has defined fixed hardware limits on the kinds of tests that can be accommodated. Ingenious software will be able to expand these limits, but will be only a partial response to this concern.

Within the timeframe of the ACAP, extensive modification of the CAT-ASVAB software is not possible without seriously disrupting the ACAP implementation timelines. Given the level of personnel resources for software development (Rafazc, 1987), the ACAP can make only a limited investment on the supplemental test battery.

These constraints point out the immediate direction of the ACAP

supplemental test battery. In the LCN, including a supplemental battery with the CAT-ASVAB should be accomplished with minimum reconfiguration of the software. Design 1 would meet this concern by adding a small number of supplemental tests and treating them as additional test modules. The drawback here is that the hardware is severely driving the selection of tests because of RAM restrictions. A strong advantage to Design 1 is that it may be accomplishable by ACAP IOT&E which will begin December 1988. Further, Design 1 could complete some of the effort needed for Design 2.

Design 2 illustrates a fully interactive configuration where the supplemental battery is just as much a part of the ACAP as the CAT-ASVAB. The design has the flexibility to present a large number of supplemental tests to satisfy both research and operational concerns (e.g., fixed test session time limits). This design would require more than a minimal amount of software development effort to accomplish. In addition, in order to have the supplemental battery and the CAT-ASVAB simultaneously resident on the TA Station, another 1 megabyte of RAM area may be required.

Design 3 represents the case where the supplemental battery is configured at the other extreme to Design 2. That is, the supplemental battery is completely independent of the testing system. Here the major advantage is that little, if any, software programming resources are required from the ACAP to generate new testing software. This does not preclude any system testing to verify that the supplemental battery is truly independent of the primary testing system. In addition, managing the multitude of disks that need to be manually handled may make logistic difficulties.

Within the stand-alone mode, Designs 3 and 4 present two scenerios. Design 3 could represent a short term remedy to meet ACAP IOT&E. Still, there is a real risk of losing response data somewhere along the pipeline as different pieces of the data reside on different media. Design 4 illustrates the case where the supplemental battery is by necessity on a separate disk, but is integrated into the testing protocol. The response data are merged at the level of the TA Station.

Summary

The ACAP testing system is required to accommodate testing on the CAT-ASVAB and on supplemental tests. This document reviewed the functional requirements for the ACAP ET Station that has only CAT-ASVAB testing. In addition, four design configurations for a supplemental test battery were described. The designs differed in the extent to which the supplemental test battery required to be integrated with the testing system that has only the CAT-ASVAB. The advantages and disadvantages of each design were briefly discussed.

References

Jones-James, G., and Rafacz, B.A.. Accelerated CAT ASVAB Project (ACAP) Development System Functional Requirements, Addendum #2: Examinee Test Station. Navy Personnel Research and Development Center, December 1985.

Rafacz, B.A., and Tiggle, R.. Functional Requirements for the Accelerated CAT ASVAB Project (ACAP) Development System. Navy Personnel Research and Development Center, November 1985.

Rafacz, B.A.. Personal communication, January 1987.

SELECTION AND CLASSIFICATION WORKING PAPER 83-4

VALIDITY OF ASVAB 8/9/10 AS PREDICTORS OF TRAINING SUCCESS

Paul G. Rossmeissl
Clessen J. Martin
Hilda Wing

U. S. Army Research Institute
for the Behavioral and Social Sciences

5001 Eisenhower Avenue
Alexandria, VA 22333

Commercial: (202-274-8275
Autovon: 8 284-8275

PP 9267

Validity of ASVAB 8/9/10 as Predictors of Training Success

Paul G. Rossmeissl
Clessen J. Martin
Hilda Wing

US Army Research Institute for the
Behavioral and Social Sciences

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple cognitive abilities test battery used by all the military services for selection and classification of enlisted personnel. Subtest scores are combined in several different ways by each military service. One combination, the Armed Forces Qualification Test (AFQT) can be linked back to assignment of military personnel during the latter days of World War II. Currently, all services use the AFQT as a primary selection hurdle for initial entry. Other combinations unique to each service were developed to predict success in training in the various military schools. The Army's combinations are termed Aptitude Area (AA) composites. Individuals wishing to join the Army must achieve minimal scores in one or more AA to be eligible to begin training in a specific Army school.

Within the past decade, however, the performance criterion favored by Army personnel managers has shifted from training performance to hands-on or job performance. This shift is a reflection of three historical trends:

A shift in training assessment from normative reference to criterion reference (i.e., Instructional System Development or ISD);

The misnorming of one set of forms of the ASVAB during 1976-1980;

The introduction of ASVAB testing into high schools for the purpose of recruiting with the "price" to the services of providing vocational counseling information to the students.

An additional issue is the attention being paid to all cognitive ability tests concerning their fairness to minority groups and to women.

Research Problem

ASVAB Forms 8/9/10 were introduced in October 1980; current plans are to have new forms replace them in October 1983. The ASVAB high school testing program is presently using (a correctly normed) ASVAB 5, but one of the current ASVAB forms will be transferred for high school use. The Army uses both training and job performance (through its Skill Qualification Test - SQT) for test validation. However, to provide timely information, it was judged necessary for all military services to provide training validity information for the preparation of the new high school Counselor's Manual and the technical manual for ASVAB 8-14. The research described below was completed to meet that requirement.

Method

Predictors

ASVAB 8/9/10 is composed of ten subtests: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto/Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). For purposes of Army selection and classification these subtests are combined into aptitude area (AA) composites (see Table 1).

- - - - - - - - - - - - - -
Insert Table 1 about here.
- - - - - - - - - - - - - -

Within the Army Composites, the AFQT is used for the initial selection of personnel and the other composites are used for the assignment of soldiers to the various MOS (military occupational specialties) or jobs within the Army.

Table 1 also shows the composition of the newly proposed ASVAB high school composites. These composites will be used for career counseling and with those students who are considering a military career. For the purposes of this research all ASVAB subtest scores were converted to standard scores. These standard scores were then combined, as in Table 1, to form the appropriate composite scores.

Criteria and Sample

The Army does not routinely record end-of-training grades on a soldier's personnel file. For this reason, during calendar year 1981 the Army Research Institute (ARI) collected training data for all MOS with 100 or more entrants per year. Included in these data were end-of-course grades for each soldier. Collection was terminated at 1000 for the high density MOS, and at the end of the year for the remaining MOS. It is these end-of-course grades which formed the criterion measures for this research. It was not possible to find useful criteria for all MOS. Many did not show sufficient variance in the end-of-course grade to be useful in the assessment of predictor validities. For example, in the MOS 16E, 92% of the grades reported were at the maximum value of 100. The analyses of this research were, therefore, limited to a sample of 1 MOS shown in Table 2. These MOS were selected because they all ad a fairly large N (operationally defined as 90 or greater) and training score standard deviation greater than five. Summary tatistics for the criterion measures from these MOS are given in able 3.

Insert Tables 2 and 3 about here.

ıalyses

The data for the MOS listed in Table 3 required further liting before any validation analyses were performed. First, :ores for all soldiers who had attrited from training for n-academic reasons were dropped. Standard scores were then mputed for those remaining. Academic attrites were assigned a ore of one standard deviation below the minimum passing score d academic recycles were assigned a score that was one-half of a andard deviation below the minimum passing score. This fferential score assignment to attrites and academic recyles s been a conventional procedure in ARI validation research volving pass/ fail training criteria and does reflect different derlying considerations between these two failure groups.

Four sets of predictors were validated against the criteria asures from each MOS: AFQT, the appropriate Army Aptitude nposite, the appropriate proposed high school composite, and the ;h school composite for Academic Ability. Uncorrected idities for these predictors were obtained using standard ;ression analyses. In addition, a stepwise regression (Draper & .th, 1966) based upon the ten ASVAB subtests was conducted for h MOS. The results of this analysis can be interpreted as the st" fit of the ASVAB subtests to the criterion data and, refore, could be used as an index for the fit of the other dictors. Validities for the composite predictors corrected for triction in range were obtained using Lawley's (1943) general e method. This method can be shown to be mathematically ntical to that proposed by Gulliksen (1950).

Finally, whenever the N within an MOS was sufficiently large to perform meaningful subgroup analyses, the above procedures were repeated for subgroups within the MOS. The variables of race and gender were used in the definition of these subgroups.

Results and Discussion

The results and discussion of this research will be divided into three topics: AA composites, subgroup analyses, and high school composites.

Operational Army Composites.

Table 4 presents the validity coefficients obtained from each of the 11 MOS for both AFQT and the appropriate AA composite. For each validity coefficient, the uncorrected and corrected value for restriction in range has been computed. Also reported for each MOS is the uncorrected stepwise best fit estimate based upon all 10 subtests of the ASVAB. Inspection of the uncorrected validity coefficients for the stepwise best fit analysis reveals that in all cases, these values are higher than for either the AFQT or for the corresponding AA composite. The average increment among the 11 MOS for the uncorrected stepwise values in comparison to the AA composite value was .10.

- - - - - - - - - - - - - - - -
Insert Table 4 about here.
- - - - - - - - - - - - - - - -

Somewhat surprising is that, for four of the 11 MOS, AFQT yielded a higher corrected validity coefficient than did the corresponding Army composite. In no instance was the increase greater than .08, as in 76P. Correspondingly, the increased predictive validity for the Army composites in relation to AFQT was greatest, in 68J and 67Y, where the increase was .11 and .09, respectively.

Inspection of the validity coefficients for the Army composites corrected for restriction in range, revealed validities ranging from .36 for 16P to .87 for 33S with an average validity coefficient of .66. The average corrected validity coefficient for AFQT was .64. The largest validities were obtained for the Skilled Technical(ST) composite (.87) and for Surveillance/ Communications(S/C) composite (.79) and the lowest average validity was for the Operators/Food (O/F) Composite (.52).

Subgroup Analyses

There were sufficient sample sizes in 16S to examine the validity coefficients separately for Blacks and Whites. Since 16S is currently not available to women because it is a combat specialty, gender comparisons were not possible. However, in 76P it was possible to examine both race and gender differences. Table 5 presents the uncorrected and corrected validities separately for Blacks and Whites in 16S for AFQT, the Operators/ Food composite, and the new high school Academic Ability composite. For the present Army O/F composite there was relatively little difference between the corrected validity coefficients for Blacks and Whites (.53 and .51, respectively). A somewhat larger difference between Blacks and Whites was observed for the AFQT (.47 and .68, respectively).

- - - - - - - - - - - - - - -
Insert Table 5 about here.
- - - - - - - - - - - - - - -

Table 6 presents the uncorrected and corrected validities in 76P, separately for Blacks and Whites as well as males and females, on AFQT, the Clerical composite, the high school Office and Supply and the high school Academic Ability composites. As in 16S, the corrected validities are somewhat higher for Whites than for Blacks and especially so for White females. However, due to the relatively small sample size for White females, this difference should be interpreted with caution. It was somewhat surprising that the Army Clerical composite had the lowest validity in comparison to the other three measures.

- - - - - - - - - - - - - - -
Insert Table 6 about here.
- - - - - - - - - - - - - - -

High School Composites

This validation research offered the opportunity to analyze the validity of the recently proposed high school composites. Table 7 presents the corrected and uncorrected validities for those MOS having an AA composite either identical to or very similar to the proposed high school composites. The high school Electronics/ Electrical composite is identical to the Army's Electronics composite, thus the results for 32D and 68J are identical to the data presented in Table 4. The high school Skilled Services composite is similar to the Army's Skilled Technical composite with the exception that AR replaces MK in the Army composite and General Science is included in the Army's composite but is not present in the high school composite. The corrected validity coefficient for the high school Skilled

Services composite was .85 and the corrected High School Academic Ability composite was .82. Inspection of Table 1 reveals that the only difference in the high school composite for Skilled Services and Academic Ability is the addition of the Mechanical Comprehension test in the Skilled Services composite; thus, the corrected validity coefficient of .82 for the Academic Ability composite was expected to approximate the validity of the Skilled Services composite for 33S.

The Mechanical Trades high school composite is similar to the Army's Mechanical Maintenance composite with the exception of Arithmetic Reasoning replacing Numerical Operations in Mechanical Trades. The resulting corrected validity in Table 7 is the same in 61B for both the Army and the high school composites. For 67Y, the high school composites was .72, compared to .75 for the Army composite displayed in Table 4.

- - - - - - - - - - - - - - - -

Insert Table 7 about here.

- - - - - - - - - - - - - - - -

The high school Office and Supply composite is similar to the Army's Clerical composite except the high school composite contains Mathematical Knowledge instead of Numerical Operations. For 71D and 76P the high school corrected validities were .67 and .69, respectively compared to the Army's Clerical composite validities of .64 and .60 as reported in Table 4. Overall, the mean corrected validity for the high school composites was .71 compared to a mean corrected validity of .64 for the Academic Ability high school composite.

The mean corrected validity of the high school composites was .71 compared to .70 for the AA composites in the seven MOS where the composites were either identical or quite similar. There were no applicable high school composites which were similar to the Army composite in the remaining four MOS.

Given the limitations, small sample size and the relatively small number of the Army's MOS analyzed, the results of this validation research were generally favorable with respect to the validity of ASVAB 8/9/10. The overall corrected validity coefficient for the Army composites was .66. In the two MOS where it was possible to analyze validities separately for Blacks and Whites, the average corrected validity coefficient for Blacks was .52 and .62 for Whites. In 76P where it was possible to compare the corrected validity coefficients for males and females, the resulting values were .61 and .58, respectively. While the overall validity of the six Army composites analyzed was .66, the average validity for AFQT across all 11 MOS was .64. This result suggests that the Army composites examined in this research

contribute relatively little in differential prediction of training criteria. This is not surprising given the limited focus of this research to training performance in a relatively few MOS. Ongoing research with a larger number of more heterogeneous MOS, using both training and job performance criteria, is anticipated to provide more definitive results. Finally, the validity of the high school composites in those MOS where they were appropriate compared very favorably to the validity of the Army composites.

Table 1

The Composition of the ASVAB Composites

Operational Army Composites

	AFQT	=	VE + AR + .5NO
Electronics	(EL)	=	AR + EI + MK + GS
Operators/Foods	(OF)	=	NO + VE + MC + AS
Surveillance/Communications	(SC)	=	NO + CS + VE + AS
Motor Maintenance	(MM)	=	NO + EI + MC + AS
Clerical	(CL)	=	NO + CS + VE
Skilled Technical	(ST)	=	VE + MK + MC + GS
Combat	(CO)	=	AR + CS + MC + AS
Field Artillery	(FA)	=	AR + CS + MC + MK
General Technical	(GT)	=	VE + AR
General Maintenance	(GM)	=	MK + EI + GS + AS

Proposed High School Composites

Mechanical Trades	=	AR + MC + AS + EI
Office and Supply	=	VE + CS + MK
Electronics/Electrical	=	AR + EI + MK + GS
Skilled Services	=	AR + VE + MC
Academic Ability	=	AR + VE

Table 2

MOS included in the Research

MOS	Name	Army Composite
05G	Signal/Security Specialist	SC
16P	Short Range Missile Crewman	OF
16S	MANPADS Crewman	OF
32D	Tech Controller	EL
33S	Electronic Warfare Systems Repairer	ST
61B	Watercraft Operator	MM
61C	Watercraft Engineer	OF
67Y	Attack Helicopter Repairer	MM
68J	Attack Fire Control Repairer	EL
71D	Legal Clerk	CL
76P	Material Control & Accounting Specialist	CL

Table 3

Summary Statistics for Training Criteria

MOS	N	Training Score Mean	Training Score S D
05G	91	84	7.3
16P	101	83	14.2
16S	514	79	8.3
32D	120	81	14.2
33S	103	82	9.0
61B	92	80	7.7
61C	150	83	6.9
67Y	137	83	6.3
68J	128	86	6.1
71D	96	73	22.9
76P	613	87	5.1

Table 4

Corrected and Uncorrected Validities
for Operational Army Composites

MOS	Uncorrected Stepwise Best Fit	AFQT Uncorrected/ Corrected	Army Composite Uncorrected/ Corrected
05G	.61	.55/.81	(SC) .48/.79
16P	.28	.15/.30	(OF) .21/.36
16S	.28	.17/.40	(OF) .23/.44
32D	.46	.44/.67	(EL) .43/.67
33S	.66	.46/.84	(ST) .56/.87
61B	.51	.49/.69	(MM) .45/.65
61C	.58	.45/.73	(OF) .45/.75
67Y	.45	.29/.66	(MM) .39/.75
68J	.53	.28/.62	(EL) .44/.73
71D	.41	.38/.65	(CL) .27/.64
76P	.48	.40/.68	(CL) .26/.60

Table 5

Validities for 16S
by Race
Uncorrected / Corrected

	Blacks	Whites
n	159	333
AFQT	.03/.47	.21/.68
O/F Composite	.16/.53	.28/.51
Academic Ability	.09/.46	.22/.48

Table 6

Validities for 76P
by Race and Gender
Uncorrected / Corrected

	Blacks	Whites
Males		
n	273	143
AFQT	.28/.69	.60/.73
CL component	.12/.57	.47/.65
Academic Ability Composite	.28/.69	.60/.74
Office and Supply Composite	.29/.67	.61/.74
Females		
n	116	38
AFQT	.26/.62	.51/.77
CL Composite	-.02/.46	.41/.69
Academic Ability Composite	.29/.65	.44/.74
Office and Supply Composite	.17/.59	.36/.71

Table 7

Uncorrected / Corrected Validities
for the Proposed High School Composites

MOS	High School Composite	Composite Validity Uncorrected/ Corrected	Academic Ability Uncorrected/ Corrected
05G	N/A	N/A	.59/.82
16P	N/A	N/A	.12/.29
16S	N/A	N/A	.17/.40
32D	Electronics/Electrical	.43/.67	.43/.67
33S	Skilled Services	.49/.85	.42/.82
61B	Mechanical Trades	.47/.65	.51/.70
61C	N/A	N/A	.46/.74
67Y	Mechanical Trades	.32/.72	.25/.64
68J	Electronics/Electrical	.44/.73	.28/.63
71D	Office and Supply	.38/.67	.39/.65
76P	Office and Supply	.42/.69	.40/.68

APPENDIX A

SAMPLE ASVAB SUBTEST MEANS AND STANDARD DEVIATIONS

	GS		AR		WK		PC		NO		CS		AS		MK		MC		EI	
MOS	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.
05G	55	6.8	56	8.0	55	5.4	56	5.0	57	5.7	56	7.2	56	8.0	55	8.0	56	7.9	55	7.7
16P	46	7.7	46	7.1	48	6.6	48	7.6	53	7.1	46	8.1	52	7.7	44	7.3	49	7.6	50	6.5
16S	47	7.4	47	7.0	46	6.9	47	7.0	52	6.6	47	6.6	53	7.0	45	5.8	51	7.0	50	6.9
32D	54	6.6	55	6.4	52	6.4	53	6.5	55	7.2	51	8.2	56	8.3	53	6.9	55	8.2	55	7.8
33s	59	6.5	60	6.0	57	4.5	57	4.9	58	5.4	55	7.4	61	5.8	58	6.6	61	5.7	61	5.8
61B	42	8.2	44	6.5	43	8.4	43	9.0	52	7.4	47	8.0	50	7.6	43	5.1	47	7.4	49	6.3
61C	53	7.2	53	6.7	51	6.8	51	7.4	53	6.7	50	7.6	61	5.2	48	6.6	57	6.0	57	5.4
67Y	55	7.2	54	7.4	53	6.5	52	6.9	54	6.3	51	7.2	61	4.9	52	8.2	60	6.1	58	5.6
68J	54	6.2	54	7.0	53	5.8	52	7.5	53	6.5	49	6.6	59	6.7	51	7.2	57	7.1	57	6.0
71D	54	8.0	57	7.6	55	6.4	55	5.8	61	4.3	60	6.2	54	8.5	54	8.6	53	9.6	53	8.1
76P	41	7.3	44	6.7	4	7.2	44	7.3	54	5.5	51	6.4	44	7.5	44	5.6	43	6.9	44	7.7

REFERENCES

Draper, N.R. & Smith, H. Applied Regression Analysis. New York, Wiley, 1966.

Gulliksen, H. Theory of Mental Tests. New york, Wiley, 1950.

Lawley, D.M. On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 1943,273-287.

DISPOSITION FORM

For use of this form, see AR 340-15; the proponent agency is TAGO.

EFERENCE OR OFFICE SYMBOL	SUBJECT
ERI-RS	Request for Validation Studies for Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10

O	FROM	DATE	CMT 1
DAPE-MPA-CS	PERI-RS		Dr. Martin/de/274-8275

. Reference DF, CMT 2, Paragraph 5, PERI-RS, ARI, 6 May 83, SAB, Inclosure 1.

:. The attached Working Paper 83-3 (Inclosure 2) entitled "Validity of ASVAB 8/9/10 as Predictors of Training Success" is submitted per The Army Research Institute's (ARI) 6 May 1983 response to DF "Request for Validation Studies for Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10" from COL Zaldo on 3 May 1983.

3. This research is based on a very limited number of Army MOS in which the sample sizes in some instances are fairly small. However, this was necessary in order to obtain MOS having training success measures with adequate psychometric characteristics.

FOR THE COMMANDER:

Newell K Eaton

for JOYCE L. SHIELDS
Director, Manpower and Personnel
Research Laboratory

2 Incl
as

PREVIOUS EDITIONS WILL BE USED

DA FORM AUG 80 2496

DISPOSITION FORM

For use of this form, see AR 340-15; the proponent agency is TAGO.

EFERENCE OR OFFICE SYMBOL	SUBJECT
APE-MPA-CS	Request for Validation Studies for Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10

O PERI-RS FROM DAPE-MPA-CS DATE 3 MAY 1983 CMT 1
Mr. Ruberton/ma/50836

. The attached request (Enclosure 1) from the Director of Personnel Plans, JSAF is forwarded for comments and information to assist in responding to MG Peek.

2. Request your comments be furnished to this office by COB 9 May 1983.

3. Response for MG Schwarzkopf, which will be prepared by this office, will include your comments concerning published and/or official validation reports for ASVAB 8/9/10 as requested in paragraph 4 of enclosure 1.

FOR THE DEPUTY CHIEF OF STAFF FOR PERSONNEL:

Enclosure

WILLIAM T. ZALDO, III
Colonel, GS
FOR: Chief, Accession Division

PERI-RS

TO: DAPE-MPA-CS FROM: PERI-RS DATE: 6 MAY 1983 CMT 2
Dr. Wing/dhl/48275

1. Reference:

a DF "Request for Validation Studies for ASVAB 8/9/10" from COL Zaldo on 28 March 1983 and our response of 19 April 1983.

b DF "Aptitude Area Composites" from COL Zaldo on 30 Nov 1982 and our response of 14 Jan 1983.

2. The Army Research Institute (ARI) is currently analyzing data collected from FY81 Army accessions, the first to be administered ASVAB 8/9/10, as part of a larger validation project. The milestones for these analyses, which are to lead to revising or verification of the Army Aptitude Area composites, are indicated in reference (b) and were the basis for the negative response to the request of reference (a). This negative response was repeated at the most recent JSSCWG meeting as noted in paragraph 3 of MG Peek's letter.

3. As stated, the criterion of interest to the Army in criterion-related validity studies is that of job performance. This necessarily implies that soldiers have to complete training and be on the job in the field for some period of time before it is appropriate to assess job performance. The index we will be using to assess the performance of soldiers in the FY81 cohort is the Skill Qualification Test (SQT). The data tapes for SQT's administered during CY 1982 -the earliest time for meaningful assessment of job performance in the FY81 cohort- are expected to be in the hands of our researchers by the end of this summer. This is the major reason why the completion of ARI analyses is not expected until October-December 1983.

CHIEF, SELECTION&CLASSIFICATION TECH AREA

U.S. GOVERNMENT PRINTING OFFICE: 1982-372-711

4. For a number of years Army training has used an Instructional Systems Development (ISD) approach which relies heavily on criterion referenced assessment of training components. When a recruit completes one component of training, s/he moves on to the next component. The outcome of this approach is that school performance criteria are frequently binary: GO/NO GO. The majority of recruits achieve a "GO" eventually, perhaps not in the initial choice of Military Occupational Specialty (MOS). Thus, training success criteria can be expected to be psychometrically weak, in that variance will be minimal. Other possible criteria such as time to completion, number of recycles, transfers, etc., are also psychometrically inadequate. For the FY81 cohort, ARI requested special information about training success in 160 MOS, specifically for criterion related validity research. Even with these special instructions for specified information, the criterion data for most of these 160 MOS were psychometrically inadequate.

5. ARI can provide partial information with respect to the criterion related validity of ASVAB as follows: There are currently two ARI technical reports in the final review and revision stages which are related to the validity of ASVAB 8/9/10. The first, a revising of the Hanser and Grafton research presented to the Defense Advisory Committee on Military Personnel Testing in the summer of 1982, is based on soldiers entering during 1976-1978 who were administered ASVAB 6/7. Analyses of the data included appropriate correction of the misnorming as well as predicting the validity of ASVAB 8/9/10, parallel to ASVAB 6/7. The criteria used were SQTs. The second by Weltin and Popelka, addresses the predictability of training success in administrative schools using the CL composite from ASVAB 8/9/10 for the FY81 cohort. If the revisions and reviews of these reports proceed as anticipated, they should be available for citation by 1 July 1983. Finally, there is a modest number of MOS in the FY81 cohort training base which have training success measures with adequate psychometric characteristics. We are preparing a brief report, by Rossmeissl and Martin, to describe what can be determined from these data concerning the validity of ASVAB 8/9/10. We expect this report to be completed, reviewed and available for citation by 1 July 1983. Nevertheless, these reports, singly and collectively, do not fully meet the Army's requirement for criterion related validity because only a portion of the criterion domain has been evaluated. Only when the research referenced in (b) is completed will the first relatively complete Army validity information be available on ASVAB 8/9/10.

6. Although no amount of forewarning could speed up the time it takes for an FY81 accession to complete SQT testing required as validation criteria, more advanced notice of such joint service requests would be very helpful in avoiding misunderstandings, research delays, or turbulence in ongoing programs. At the recent Tri-Service Lab Commanders and Technical Directors' Conference at West Point Apr 25-27, this topic was addressed. The group

ERI-RS (6 May 83)
UBJECT: Request for Validation Studies for Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10

greed on the utility of projecting, prior to each FY, as many of the next ear's requirements as possible. Dr. Sicilia, Director of Accession Policy OASD-MRA&L), agreed to take the lead in guiding the JSSCWG and related orking groups toward that end. We anticipate these actions will enhance our bility to provide timely and useful support to DCSPER and the Joint-Service roups.

OR THE COMMANDER:

2 Incl.
1. DF, 28 Mar 83
2. DF, 14 Jan 83

JOYCE L. SHIELDS
Director, Manpower and Personnel Research Laboratory

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS UNITED STATES AIR FORCE
WASHINGTON, D.C. 20330

28 APR 1983

REPLY TO ATTN OF: MPX

SUBJECT: Request for Validation Studies for Armed Services Vocational Aptitude Battery (ASVAB) 8/9/10

TO: Major General H. N. Schwarzkopf

1. ASVAB 14 is scheduled to replace the current high school test on 1 July 1984. Two Air Force contracts, development of a new Counselor's Manual and use of the Validity Generalization concept to relate ASVAB 14 to military and civilian occupations, are critical to timely and successful test implementation. A third contract, development of a technical manual for ASVABs 8 through 14, while not critical for test implementation, will provide important support for the technical credibility of all ASVAB versions.

2. For the past three fiscal years, the services have been provided Joint-Service ASVAB Research and Development funds to conduct studies that validate the relationship between ASVAB test scores and training performance criteria. These studies are required to support the three contracts mentioned earlier and, perhaps more importantly, to clearly demonstrate the predictive utility of our enlistment tests.

3. Attached is a 16 March 1983 letter from the Chairman of the Joint-Service Selection and Classification Working Group (JSSCWG, formerly ASVAB Working Group), which requests the services provide published technical validation reports to the Air Force Human Resources Laboratory (AFHRL) by 1 May 1983. At the 18-22 April 83 JSSCWG meeting, service technical representatives indicated they could not meet this request and in one case would not have the required data until Oct 1983.

4. To meet the ASVAB 14 implementation schedule and provide a firm technical foundation for all ASVABs, we require all published and/or official validation reports for ASVABs 8/9/10 no later than 1 July 1983. We would appreciate your strong support of this requirement. It is our understanding that validity studies will be an agenda item at our next Steering Committee meeting on 11 May. Relevant reports should be provided to Major John Welsh, Chairman, JSSCWG, AFMPC/MPCYPT, Randolph AFB, TX 78150. My action officer is Lt Col Jim Watson, MPXOA, AUTOVON 225-4679.

Kenneth L. Peek, Jr.

KENNETH L. PEEK, JR.
Major General, USAF
Director, Personnel Plans

1 Atch
AFMPC/MPCYPT Ltr, 16 Mar 83

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS AIR FORCE MANPOWER AND PERSONNEL CENTER
RANDOLPH AIR FORCE BASE, TX 78150

16 MAR 1983

REPLY TO ATTN OF: MPCYPT

SUBJECT: Request for Validation Studies for ASVAB 8/9/10

TO: ASVAB Working Group Members - Testing Policy Representatives

1. The Air Force Human Resources Laboratory (AFHRL) has recently let a contract with McFann-Gray Associates Inc. to develop a technical manual that will support ASVABs 8 through 14. In order to complete this effort in a timely manner, AFHRL needs to have all published technical validation reports that address the validities of ASVABs 8/9/10. Request this material be provided to Dr. Malcolm Ree, AFHRL/MOA, Brooks AFB, Texas, 78235, by 1 May 83.

2. I am aware that all the Services have these necessary validation studies in various stages of completion. In the event that you cannot provide written reports of the studies, please let me know when such reports can be made available, along with a negative reply to the 1 May 83 suspense. The topic of ASVABs 8/9/10 validity studies will be an agenda item at the 18-19 Apr 83 Working Group meeting.

JOHN R. WELSH, Maj, USAF
Chairman, ASVAB Working Group

Cy to: OASD(MRA&L), Dr. Sellman

AFHRL/MO, Lt Col Amor

McFann-Gray Associates Inc., Dr. Shore

DISPOSITION FORM

For use of this form, see AR 340-15; the proponent agency is TAGO.

REFERENCE OR OFFICE SYMBOL	SUBJECT
DAPE-MPA-CS	Request for Validation Studies for ASVAB 8/9/10

TO PERI-RS FROM DAPE-MPA-CS DATE 28 MAR 1983 CMT 1
Mr. Ruberton/njd/50836

1. Request comments/copies of published reports (as requested by the attached letter from the Air Force) be provided to this office by 22 April 1983.

2. If the reports are forwarded direct to the Air Force, please furnish a copy of your transmittal document to this office.

FOR THE DEPUTY CHIEF OF STAFF FOR PERSONNEL:

WILLIAM T. ZAIDO, III
FR: Colonel, GS
Chief, Accession Division

1 Enclosure

TO Chief, Accession Division FROM PERI-RS DATE 19 APR 1983 CMT 2
ATTN: DAPE-MPA-CS C. Martin/jld/274-8275

1. ARI will not have any written reports on the validation of ASVAB 8/9/10 by 1 May 83.

2. We expect to have some validation reports in December,1983 as a result of Task 1 activities in Project A.

FOR THE COMMANDER:

signed

JOYCE L. SHIELDS
Director, Manpower and Personnel
Research Laboratory

1 Incl
nc

1 Incl.

DA FORM AUG 80 2496 PREVIOUS EDITIONS WILL BE USED U.S. GOVERNMENT PRINTING OFFICE: 1982-372-711

READING FILE/SELECTION&CLASSIFICATION TECH AREA

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS AIR FORCE MANPOWER AND PERSONNEL CENTER
RANDOLPH AIR FORCE BASE, TX 78150

16 MAR 1983

REPLY TO ATTN OF: MPCYPT

SUBJECT: Request for Validation Studies for ASVAB 8/9/10

TO: ASVAB Working Group Members - Testing Policy Representatives

1. The Air Force Human Resources Laboratory (AFHRL) has recently let a contract with McFann-Gray Associates Inc. to develop a technical manual that will support ASVABs 8 through 14. In order to complete this effort in a timely manner, AFHRL needs to have all published technical validation reports that address the validities of ASVABs 8/9/10. Request this material be provided to Dr. Malcolm Ree, AFHRL/MOA, Brooks AFB, Texas, 78235, by 1 May 83.

2. I am aware that all the Services have these necessary validation studies in various stages of completion. In the event that you cannot provide written reports of the studies, please let me know when such reports can be made available, along with a negative reply to the 1 May 83 suspense. The topic of ASVABs 8/9/10 validity studies will be an agenda item at the 18-19 Apr 83 Working Group meeting.

JOHN R. WELSH, Maj, USAF
Chairman, ASVAB Working Group

Cy to: OASD(MRA&L), Dr. Sellman

AFHRL/MO, Lt Col Amor

McFann-Gray Associates Inc., Dr. Shore

PERI-RS (30 Nov 82)
SUBJECT: Aptitude Area Composites

TO Chief, Accession Division DAPE-MPA-CS

FROM PERI-RS

DATE 14 Jan 83 CMT 2
Dr. Mitchell/drb/48275

1. The Army Research Institute is currently involved in a major research effort directed at the development and validation of Army selection and classification measures. The research is designed to improve existing and/or develop new preinduction predictors of success in Army school training, Army-wide success, and performance of specific MOS-related duties. An evaluation of the aptitude area composites is part of this effort.

2. Data necessary to evaluate the validity and utility of the operational Army aptitude area (AA) composites include (1) ASVAB 8/9/10 data for Army applicants and accessions, (2) training grades or training performance ratings for accessions, and (3) SQT and other job performance measurements for recruits. While data in (1) are currently available for many FY81 and FY82 applicants, current data in (2) are limited to end-of-course grades for some FY81 accessions in selected MOS. SQT data (3) are not now available for most FY81 recruits; they will be available for the FY81 cohort during this fiscal year. Other performance data currently available are attrition, promotion, and disciplinary data for a limited number of FY81 accessions, but other specific job performance measures are generally unavailable. The data described at (1), (2), and (3) are expected to be essentially complete by 1 October 1983. Appropriate analyses to develop, verify, and document an acceptable revision of aptitude area composites will require 6-9 months, followed by a 3-month professional and legal review. Thus, it is not feasible to meet the 1 October 1983 deadline suggested above.

3. A more extensive set of data will be collected from a sample of the FY83/84 cohort. This will include a broader variety of predictor measures as well as performance measures, and is expected to be compiled by 30 September 1985. These analyses will focus on (1) the reliability of composite scores over various ASVAB forms and administrations, (2) the predictive validity of composite scores with respect to performance in training (as measured by appropriate level and time-to-master indices) separately for high density MOS, (3) the predictive validity of composite scores with respect to on-the-job performance (as measured by ratings and hands-on performance measures indexed to high frequency and representative job tasks) separately for high-density MOS, (4) the generalizability of validity coefficients to training and job performance for MOS not examined, (5) the differential validity of the aptitude area composites with respect to training/job performance for individual MOS, and (6) the utility, in terms of improved performance, of selection/classification decisions based on the additional information provided by an increasingly greater number of AA composites. Preinduction predictor data for performance in Army school training, Army-wide success, and performance of specific MOS-related duties will be used to verify and extend the aptitude area composite set developed during FY84 as described in paragraph 2.

2 Incl.

PERI-RS (30Nov 82)
SUBJECT: Aptitude Area Composites

4. In order to meet the Army's need for re-evaluated ASVAB aptitude area scores with appropriate psychometric/legal documentation, ARI proposes to examine ASVAB 8/9/10, training, and SQT data for the F81 cohort. A research plan for preliminary examination of the AA composites will be based on FY81 cohort data. The plan will be developed by May 1983. We will coordinate the results of this examination with your office as analyses are completed between this May and next winter. We expect the preliminary evaluation of the AA composites to be complete by 15 March 1984. The final evaluation should be complete by 15 May 1984. The system should be available for DCSPER review 1 August 1984. Implementation will occur pending approval by DCSPER. Request this date be used as a milestone for changing the AA composites. More extensive analyses, based on a much more complete data set for FY83/84 recruits, and yielding likely additional alterations to AA composite structure, should be available for DCSPER review by 1 August 1986.

FOR THE COMMANDER:

for JOYCE L. SHIELDS, Ph.D.
Director, Manpower and
Personnel Research Laboratory

DISPOSITION FORM

For use of this form, see AR 340-15; the proponent agency is TAGO.

REFERENCE OR OFFICE SYMBOL	SUBJECT
DAPE-MPA-CS	Aptitude Area Composites

TO PERI FROM DAPE-MPA DATE 30 NOV 1982 CMT 1
Mr. Ruberton/njd/50836

1. At present the Army uses ten different Aptitude Area Composite scores (AA) based on the Armed Services Vocational Aptitude Battery (ASVAB) for enlistment purposes. The ASVAB subtest composition of these AA is linked to the Army Classification Battery (ACB-73) which was used prior to implementation of ASVAB in 1976.

2. There are several issues involved with the continued use of ten Army AA composites.

a. Both ACB-73 and ASVAB 6/7 (used 1 Jan 76 - 30 Sep 80) had 16 subtests while ASVAB 8/9/10 (used since 1 Oct 80) and projected follow-on forms have ten subtests. The current AA overlap in content and are highly intercorrelated.

b. The criterion for AA training success in Army schools, reflected traditional courses based on lectures and demonstrations. Some current Army training methods emphasize self-paced learning and practical exercises.

c. In recent years the criterion of job performance has been emphasized.

d. Many of the occupational specialties in the Army are the same as or similar to occupations in the other military services. Such occupations (e.g., clerical/administrative, mechanical, electrical) should have the same AA composites since training in many MOS is accomplished by a single Service.

3. Decisions about changing the number and composition of Army AA need to be made in the near future.

a. Adoption of the 1980 Profile of American Youth Population as the reference population is being considered by OSD.

b. New forms of ASVAB are scheduled for implementation o/a 1 October 1983.

4. Request evaluation of the feasibility of changing the Army Aptitude Area Composites by 1 October 1983. If this is not feasible, request the length of time required for evaluating the composites.

5. Your comments by 3 January 1983 will be appreciated.

FOR THE DEPUTY CHIEF OF STAFF FOR PERSONNEL:

George W. Kemp CPT
for WILLIAM T. ZALDO, III
Colonel, GS
Chief, Accession Division

DA FORM AUG 80 2496 PREVIOUS EDITIONS WILL BE USED

★U.S. GOVERNMENT PRINTING OFFICE: 1982-372-711

Working Paper

RS-WP-86-04

PROJECT A DATA BASE SECURITY: AN UPDATE AND A REMINDER

Paul G. Rossmeissl and Lauress L. Wise

March 1986

REVIEWED BY:
JANE M. ARABIAN
Research Psychologist,
SCTA

APPROVED BY:
LAWRENCE M. HANSER
Acting Chief, Selection and
Classification Tech Area

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

PROJECT A DATA BASE SECURITY: AN UPDATE AND A REMINDER

PAUL G. ROSSMEISSL
AND
LAURESS L. WISE

THE NEED FOR SECURITY

Whenever a large amount of data on individuals is maintained and stored, it is necessary to develop procedures to protect that data from compromise. The security of the Project A and B LRDB is particularly important for a number of reasons. Some of the data collected on individual soldiers, such as promotions, paygrade, or disciplinary actions, will be private in nature, and the privacy of that information must be maintained. Since many researchers will be accessing the LRDB for a variety of uses, the integrity of the data must be maintained to ensure that the data remain accurate and consistent across uses. Finally, it is necessary to secure the database to ensure that the Army maintains ownership of the data. In other words, to ensure that the data within the LRDB are used only for authorized Project A and B research.

SECURITY PROCEDURES

The security of the LRDB will be protected in three ways. Soldier social security numbers (SSNs) will be routinely encrypted to ensure the privacy of each soldier's records. Access to the LRDB will be controlled both to further protect soldier privacy and to ensure proper use of the data. Finally, a log will be maintained for the LRDB system that will note each attempted access of the LRDB and whether the access was authorized or not. Each of these procedures will be detailed in the subsections that follow.

SSN ENCRYPTION

The key aspect to guaranteeing the privacy of individual soldier data will be the coding or encrypting of each soldier's identifier. This encryption will be accomplished by scrambling each soldier's SSN in an unpredictable way. The algorithm that will do the encrypting (and if needed, decrypting) will be known only to the LRDB manager and ARI in-house database managers. A printed copy of the algorithm will be securely maintained by the Project A data base manager. All of the data files of the LRDB that can be routinely accessed and any project workfiles generated from the larger LRDB will use only this encrypted SSN as the soldier identifier.

CONTROLLED FILE ACCESS

The integrity and accuracy of the LRDB data will be maintained by controlling the access to the large files or relations within the database. This procedure will also further contribute to the privacy protection of individual soldier records. In general, the system to be adopted will use the RACF procedures available at NIH to allow the access of particular files to authorized users. Under RACF, different levels of access can be granted to different users. By specifying a "universal access" of "NONE" on all of Project A datafiles, access will be restricted to only those users granted specific exceptions. In most cases, such users will be allowed "READ" access only. Such users will have to provide an eight-character RACF password (different for each user) in order to read the datafiles for which they have been given access. Using the provisions of RACF, a series of access "levels" will be developed which should provide timely access to relevant data needed by Project A and B researchers and yet protect the security and integrity of the data.

Level 1. At the highest level of access will be the database managers. Currently, these individuals are Dr. Lauress Wise and Ms. Winnie Young of AIR and Dr. Paul Rossmeissl of ARI. Level 1 personnel will have access to all of the files within the database. Furthermore, only Level 1 personnel will be able to officially enter data into the database or modify data already stored in the database. Thus, the database managers must assume responsibility for data entry, editing, and the storage of original data materials (i.e., tapes, diskettes) in a secure location. In addition, it will be the duty of the Level 1 personnel to create Level 4 workfiles as they are needed by other project researchers.

Level 2. Personnel at the second access level will be able to directly read data from all of the files in the database with the exception of the Link File (see LRDB Plan Section 4.3.2.) that contains basic soldier identifying information. This exception is made to maintain soldier privacy. Dr. Lawrence Hanser, the current COTR will also have Level 2 access as will Ms. Frances Grafton of ARI.

Level 3. Most project personnel will have some Level 3 LRDB access. Researchers at this level will have direct access to all files that are generated by the particular tasks they are investigating. Furthermore, they will have direct access to the files created by other tasks that directly impact their work. For example, Task 2 researchers will have direct access to the task analysis data collected by Task 5 so that the new predictors that are developed will address areas of the criterion space not currently covered by ASVAB.

Level 4. The most common way in which project researchers will access the LRDB is through the creation of workfiles (see LRDB Plan Section 4.4.). By requesting the creation of a workfile, a researcher will be able to obtain data from all of the large files in the database except the Link File which will contain soldier identifying information and will always be kept private and secure. The key aspect of workfiles relevant to LRDB security is that the researcher will only receive the data that he or she requested and there will be a precise record of who requested that data. When a project scientist requires a workfile, he or she will submit a data request form to either the contractor or ARI database manager. This request form will ask:

(1) Who wants the data?

(2) What variables are needed?

(3) What sample is needed?

(4) Which LRDB file or files contain the data being requested?

(5) Why are the data needed?

(6) Will the data be downloaded to hardware other than the NIH computer facility?

(7) What will be done with the data after its current use is completed (i.e., file will be scratched or saved for future use)?

Requests from Project A staff for workfiles can be initiated by typing the WYLBUR command "WORKFILE". This command will present the user with a workfile request form, that can then be WYLBUR mailed to the data base managers (initials LWV, WYO, or GAR). Such requests will be reviewed by the LRDB Manager and the Task 1 Monitor for appropriateness. If any dispute should arise, the COTR, Principal Scientist, and relevant Task Leader will confer with the LRDB Manager and Task 1 Monitor to resolve the dispute.

Requests from researchers outside of Project A for workfiles will be reviewed by the ARI Task 1 Monitor and the COTR for approval. Release of data outside of Project A but within ARI is an ARI policy issue. Requests for data files originating outside of ARI must be approved by ODCSPER as required by the Project A/B LOA with that office.

Whenever workfiles are created, the researcher(s) receiving the files will agree to the following:

(1) These data are the property of the Army and the researcher accepts responsibility for preventing further release of these data without formal ARI approval. This means, in

particular, that derivative files with records on individual soldiers will be protected by RACF or an equivalent system.

(2) Papers, publications, and briefing charts based upon these data will be submitted to ARI for clearance before being shown outside of Project A.

(3) The researcher will inform the LRDB Manager when analyses of the workfile have been completed so that it can be purged.

LRDB ACCESS LOG

A report will be generated each week showing the number of accesses to each LRDB datafile, including each workfile, by each authorized user. In addition, the monthly accounting information of each project user will be monitored for any indication of unauthorized access to the LRDB. These audit trials will serve as a second level of protection against unauthorized use of the data by anyone who manages to obtain the necessary RACF passwords. They will not directly prevent unauthorized access to the LRDB, but the threat of exposure should serve as a significant deterrent to attempts to unauthorized LRDB access. The log will also help the database managers decide which project files should be stored on disk rather than tape by providing information as to how frequently data are requested from any give file.

OTHER SECURITY PROCEDURES

Much of the data that will be entered into the LRDB will come from existing Army sources, such as the EMF. Additional precautions beyond those mentioned above will be taken to secure the information on these data tapes. The key aspect of this additional security is to collect and store information from these sources only if it is essential to the goals of Project A and B. For example, with regard to the EMF, this LRDB plan indicates specifically which variables will be needed. Other variables, in particular, each soldier's location and assigned unit, will not be routinely acquired and stored. In addition to limiting the data elements to be stored, the number of soldiers for whom any data will be retained will be limited. As indicated in Section 3 above, the LRDB will not obtain and keep information on all active service personnel. Only data from personnel selected for Project A and B research will be maintained.

SUMMARY

Any set of procedures designed to store data electronically needs to balance the ease with which data can be accessed against the security of the database. The procedures presented in this section tend to favor the security aspect of this balance. The number of data files that most project researchers will be able to access directly will be quite limited. Furthermore, only the

database managers will be able to add or modify data, and access the true soldier identifying information. However, efficient use and rapid creation of the workfiles should provide any project scientist with the data that he or she needs to perform the required research.

REFERENCE

Wise, L. L., Wang, M., and Rossmeissl, P. G. Development and Validation of Army Selection and Classification Measures Project A: Longitudinal Research Database Plan. Research Report # 1356, U. S. Army Research Institute, December 1983.

Project A Workfile Request Form

Name ____________________________________

NIH RACF ID ______________________________

Information on the Data that you need

1. From which LRDB file or data collection do you need data?

2. Which variables from the total data set are needed?

3. What sample or subsample of the cases is needed?

4. Why are these data needed? (Please attach a short outline of the planned research.)

5. Will the data be downloaded to any system other than the NIH computer facility?

6. What will be done with the data after its current use is completed (i.e., the file will be scratched or saved for future use)?

Please note that as you request a workfile you agree to the following policy:

(1) These data are the property of the Army and the researcher accepts responsibility for preventing further release of these data without formal ARI approval. This means, in particular, that derivative files with records on individual soldiers will be protected by RACF or an equivalent system.

(2) Papers, publications, and briefing charts based upon these data will be submitted to ARI for clearance before being shown outside of Project A.

(3) The researcher will inform the LRDB Manager when analyses of the workfile have been completed so that it can be purged.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-3

Project A Proponent Feedback Briefing Reviewing Status and Results for MOS 63B: June 18, 1987

Michael G. Rumsey and Darlene M. Olson
Selection and Classification Technical Area

Deirdre J. Knapp
Human Resources Research Organization

January 1989

REVIEWED BY:

Jane M. Arabian
Acting Team Chief, Test
Validation and Standards, SCTA

APPROVED BY:

Frances C. Grafton
Acting Chief, Selection and
Classification Technical Area

CLEARED BY:

for Newell K. Eaton, Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

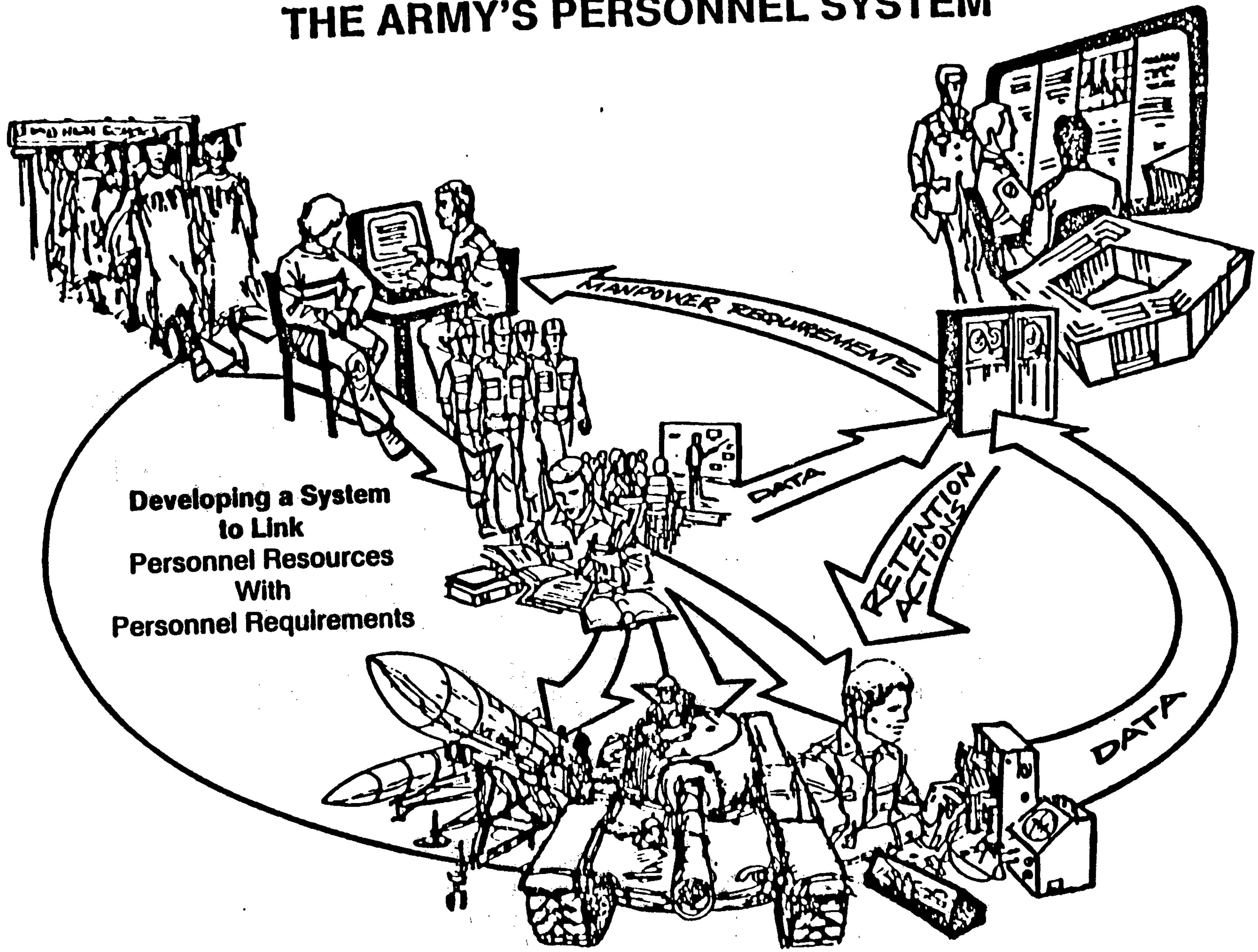
PROJECT A PROPONENT FEEDBACK BRIEFING
REVIEWING STATUS AND RESULTS FOR MOS 63B:
JUNE 18, 1987

Project A is a major component of a large-scale effort to improve the Army's selection and classification system. In 1985, a major Project A data collection was completed involving over 9400 soldiers. Following the completion of the principal set of analyses on these data, briefings were offered to Army agencies who had provided support to the research and might have some interest in the results obtained. This working paper, prepared to meet a specific Army request for information, presents slides and accompanying narrative for one such briefing, presented by an Army Research Institute representative, Michael Rumsey, to proponents for the Light Wheel Vehicle Mechanic (63B) military occupational specialty and other attendees at Aberdeen Proving Ground, Maryland, on June 18, 1987. Where the information presented is no longer current, explanatory notes have been inserted.

PROJECT A:

A REVIEW OF CURRENT STATUS AND RESULTS FOR MOS 63B

THE ARMY'S PERSONNEL SYSTEM



PROJECT A
RESEARCH OBJECTIVES

- Validate ASVAB Against Job Performance *
- Develop and Validate New and Other Existing Selection and Classification Procedures and Measures to Optimize the Person-Job Match
- Develop Computer-Based Decision-Aids for Managers of the Army's Manpower Processes

* Congressional Mandate

PROJECT A MANAGEMENT GROUP

Past General Officer Advisors

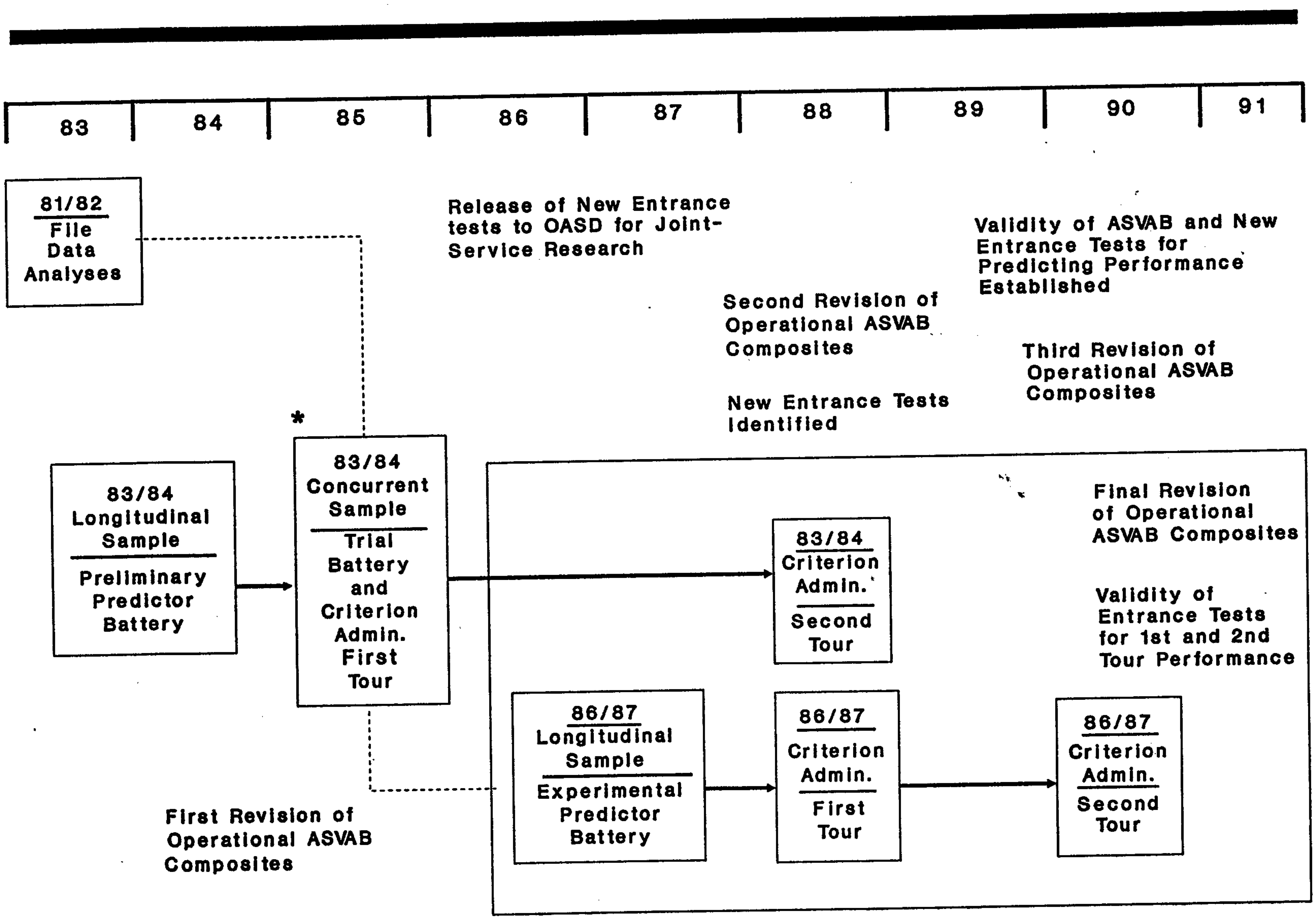
MG C.F. BRIGGS
MG R. BROWN
MG J.H. CORNS
MG M.O. EDMONDS
MG J.W. FOSS
MG F.M. FRANKS, JR.
BG W.C. KNUDSON
BG G.E. LUCK
MG B.B. PORTER
MG N. SCHWARZTKOPF

Current Governance Advisory Group

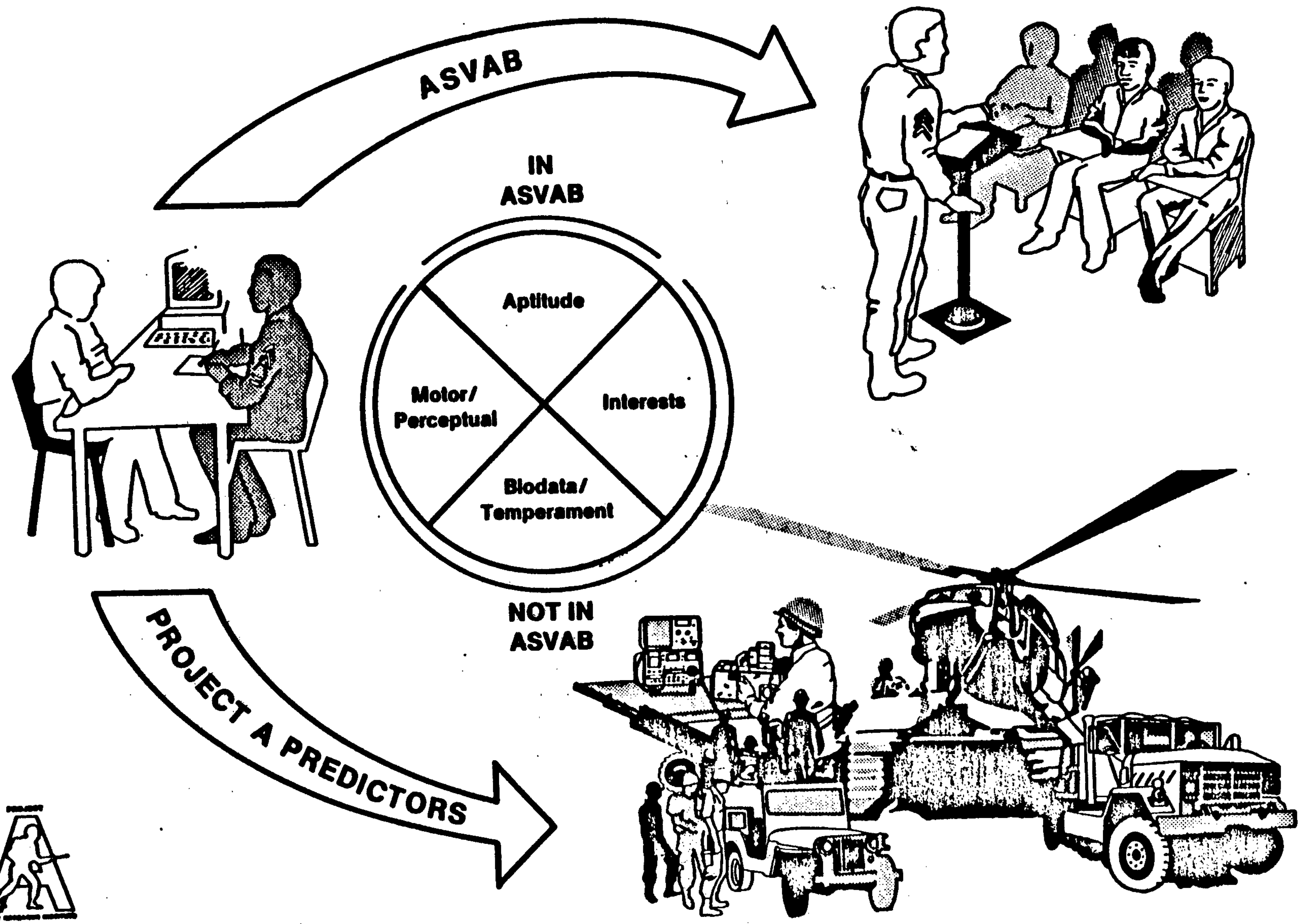
MG W.G. O'LEKSY
CHAIRMAN

SCIENTIFIC ADVISORS	U.S. ARMY ADVISORS	INTERSERVICE ADVISORS
DR. P. BOBKO	MG J.B. ALLEN, JR.	DR. N.K. EATON (Chair)
DR. T. COOK	MG T.J.P. JONES	DR. N. GUINN (AF)
DR. M. HAKEL (Chair)	BG G. MALLORY	DR. W.S. SELLMAN (DoD)
DR. L. HUMPHREYS	BG P. MALLORY	
DR. L. JOHNSON		
DR. R. LINN		
DR. M. TENOPYR		
DR. J. UHLANER		

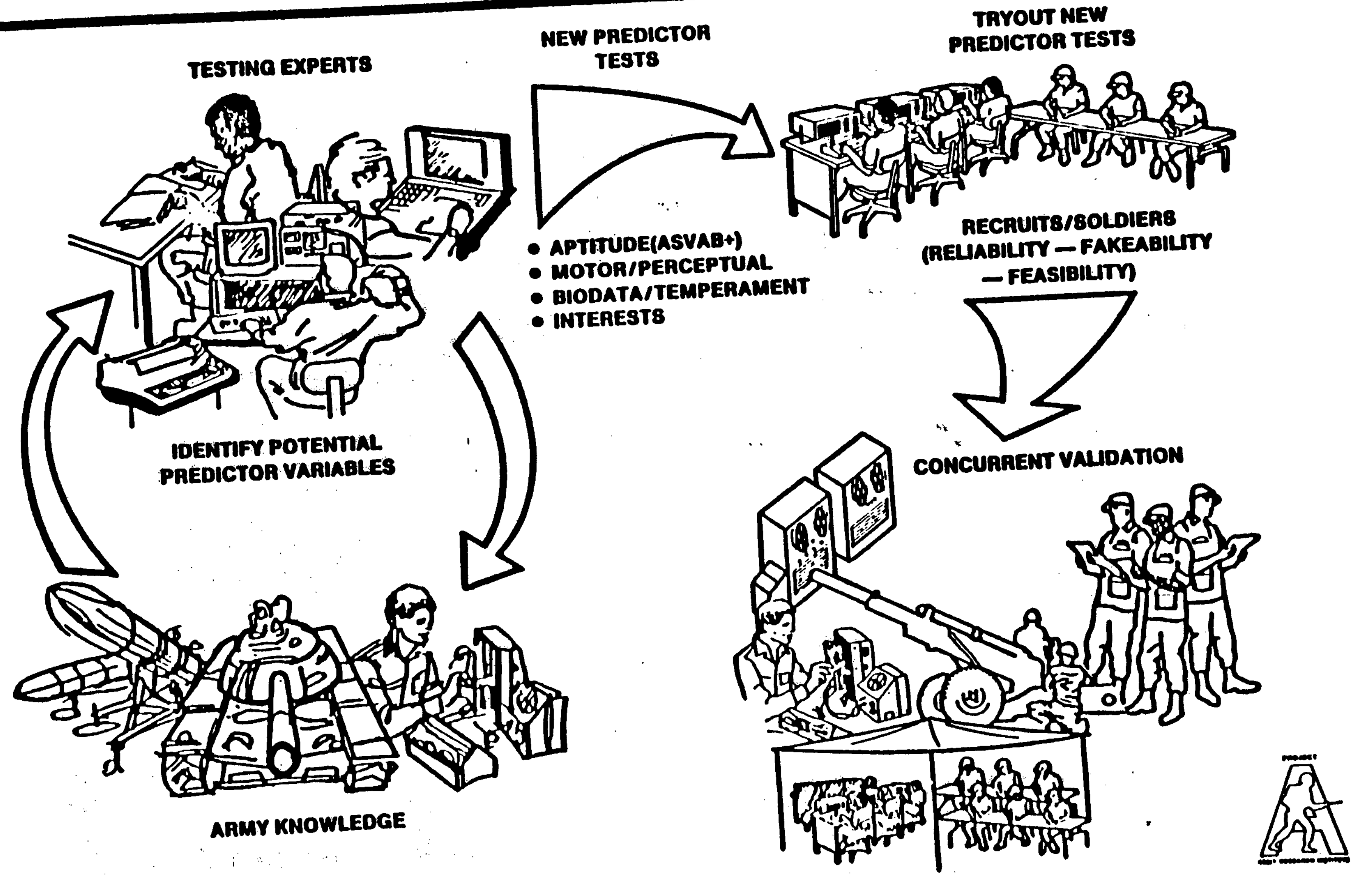
PROJECT A SCHEDULE AND MAJOR PRODUCTS



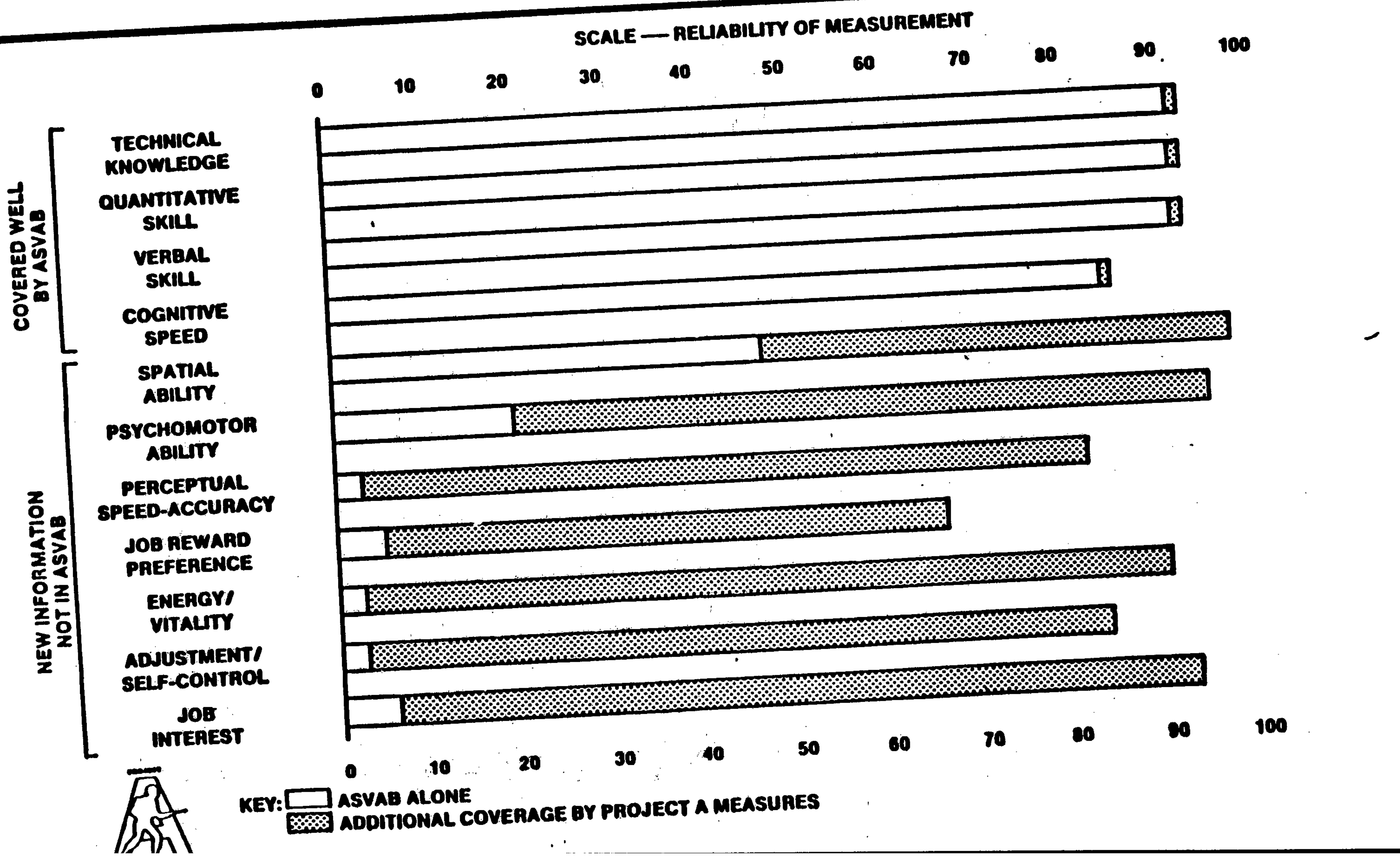
PROJECT A PREDICTORS



STRATEGY FOR DEVELOPING NEW PREDICTOR MEASURES



COVERAGE OF PREDICTOR DIMENSIONS ASVAB VS NEW TESTS



ESTABLISH ARMY CONSENSUS ON MEASURING JOB PERFORMANCE

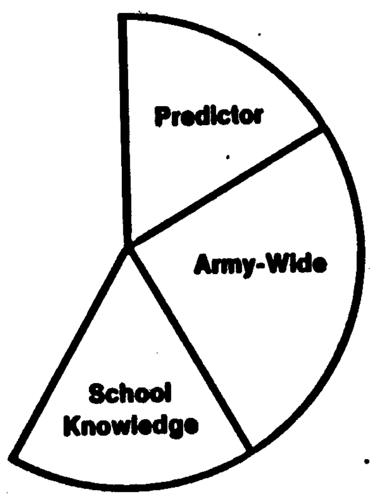
GOAL:

- Reflect Military Judgment About What it Means to be a Good Soldier
- Be Consistent With Written Doctrine
- Be Consistent With Proponent Philosophy
- Be Consistent With Field Practice

PROJECT A PERFORMANCE MEASURES

FULL TREATMENT
(Batch A)

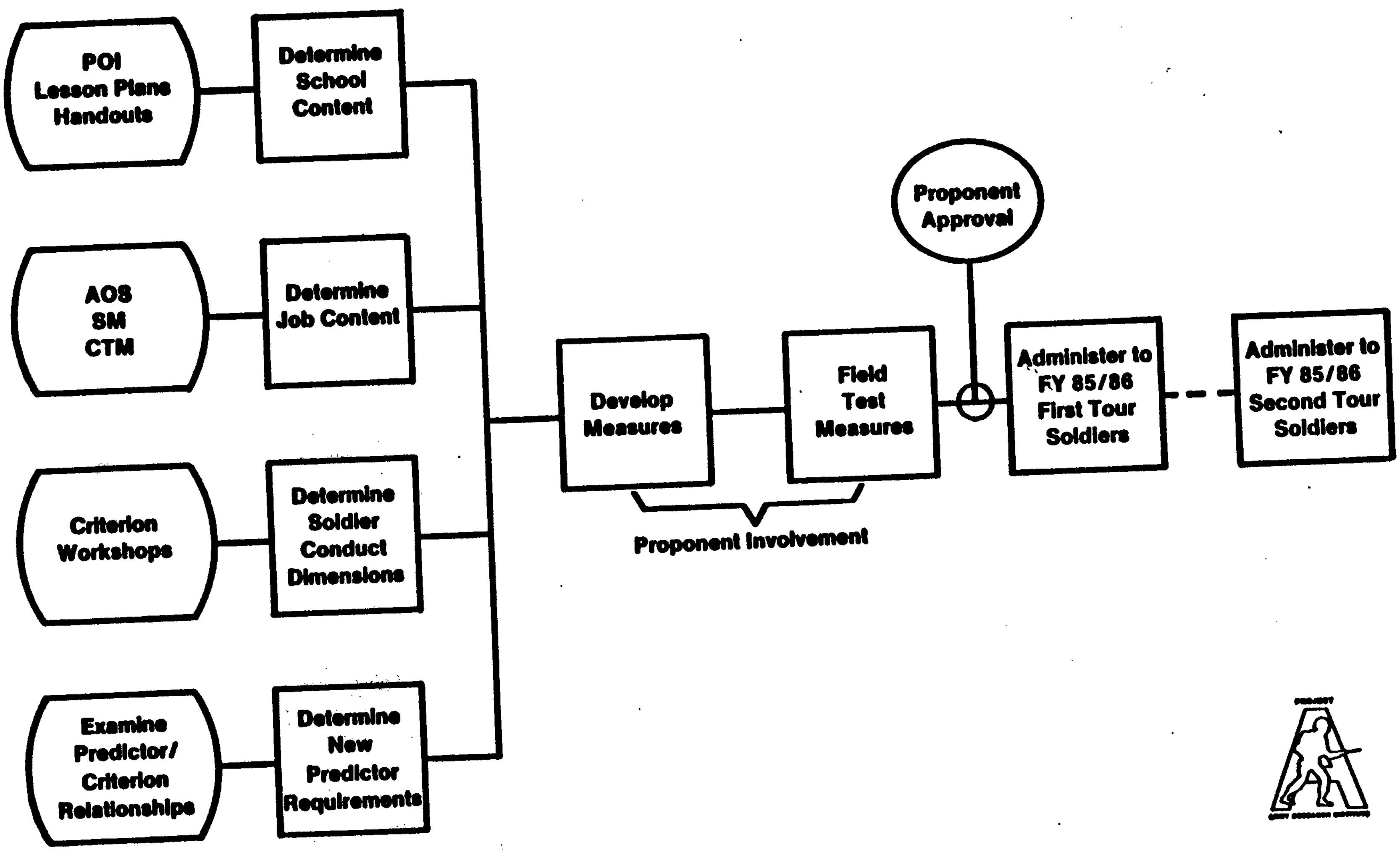
MOS	TITLE	MOS	TITLE
13B	Cannon Crewman	11B	Infantryman
64C	Motor Transport Oper	19E	Tank Crewman
71L	Admin Specialist	31C	Radio TT Oper
95B	Military Police	63B	Vehicle & Generator Mech
		91A	Medical Care Specialist



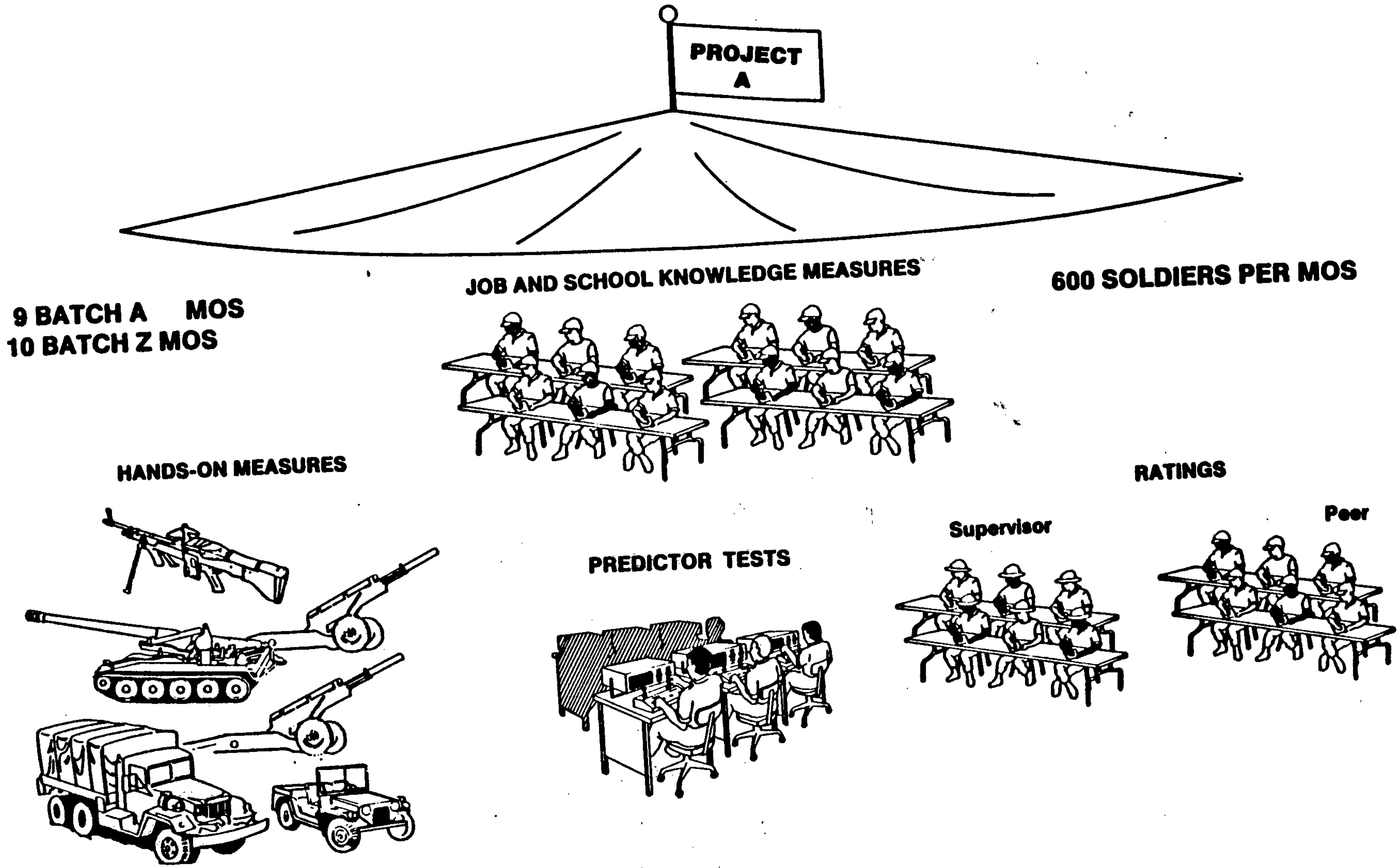
PARTIAL TREATMENT
(Batch Z)

MOS	TITLE	MOS	TITLE
12B	Combat Engineer	55B	Ammunition Spec
16S	MANPADS Crewman	67N	Utility Helicopter Rpr
27E	Tow/Dragon Rpr	76W	Petroleum Supply Spec
51B	Carpentry/Masonry Spec	76Y	Unit Supply Spec
54E	Chemical Operations Spec	94B	Food Service Spec

DEVELOPING NEW MEASURES



CONCURRENT VALIDATION



CONCURRENT VALIDATION SAMPLE SOLDIERS BY MOS BY LOCATION

Location \ MOS	BATCH A 11B	13B	19E	31C	63B	64C	71L	91A	95B		BATCH Z 12B	16S	27E	51B	54E	55B	67N	76W	76Y	94B	Total	% Total
Ft. Benning	45	23	41	7	13	39	16	9	13		13	15	3	0	12	18	9	13	15	12	316	3.35
Ft. Bliss	0	20	30	15	61	45	17	0	44		15	5	2	0	14	0	12	6	31	30	347	3.68
Ft. Bragg	68	46	0	0	37	25	41	10	72		82	75	13	19	72	20	7	42	39	62	730	7.74
Ft. Campbell	90	28	0	20	60	45	54	44	43		90	23	10	0	32	18	42	51	61	46	757	8.03
Ft. Carson	60	50	77	30	49	53	30	33	46		49	57	13	0	25	7	0	23	40	47	689	7.31
Ft. Hood	26	56	0	30	40	28	38	50	60		51	60	4	12	82	36	44	72	41	57	767	8.13
Ft. Knox	29	32	111	16	38	48	22	45	31		43	10	6	0	8	12	0	10	29	34	524	5.56
Ft. Lewis	75	46	13	11	43	46	23	27	56		27	25	1	11	51	31	20	48	41	36	631	6.69
Ft. Ord	30	0	0	14	30	42	31	43	51		51	7	8	1	4	7	15	23	40	28	425	4.51
Ft. Polk	73	47	19	29	47	47	18	46	44		60	45	9	8	16	7	23	26	51	35	648	6.87
Ft. Riley	30	43	55	27	26	45	35	30	40		31	20	8	8	25	52	0	20	39	45	579	6.14
Ft. Sill	0	108	0	20	43	51	44	0	29		42	11	0	0	0	0	15	7	35	32	437	4.63
Ft. Stewart	44	46	39	17	28	51	31	45	45		30	39	9	8	17	29	26	44	34	35	617	6.54
USAREUR	132	122	120	130	122	121	114	119	118		120	78	61	41	96	54	63	105	134	113	1963	20.8
Total	702	667	503	366	637	686	514	501	692		704	470	147	108	434	291	276	490	630	612	9430	
% Total	7.44	7.07	5.33	3.88	6.76	7.27	5.45	5.31	7.34		7.47	4.90	1.56	1.15	4.60	3.09	2.93	5.20	6.68	6.49		

PROJECT A
RESULTS FROM THE CONCURRENT VALIDATION

- Description of Job Performance Levels
- Prediction of Job Performance Using the ASVAB
- Prediction of Job Performance Using the New Project A Predictors

MEASUREMENT METHODS

Hands-On MOS Specific Task Tests
Written MOS Specific Task Tests
Supervisor Ratings of Technical Skill

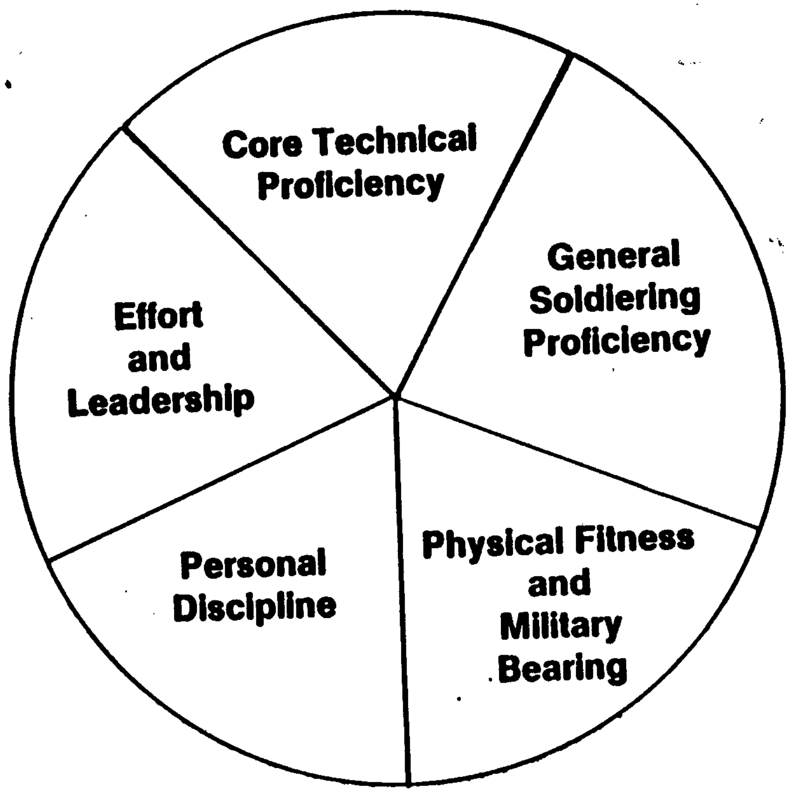
Hands-On Tests of Common Soldier Tasks
Written Tests of Common Soldier Tasks

Ratings of: Effort/Leadership
Self-Development
Awards and Certificates
Combat Effectiveness Appraisals

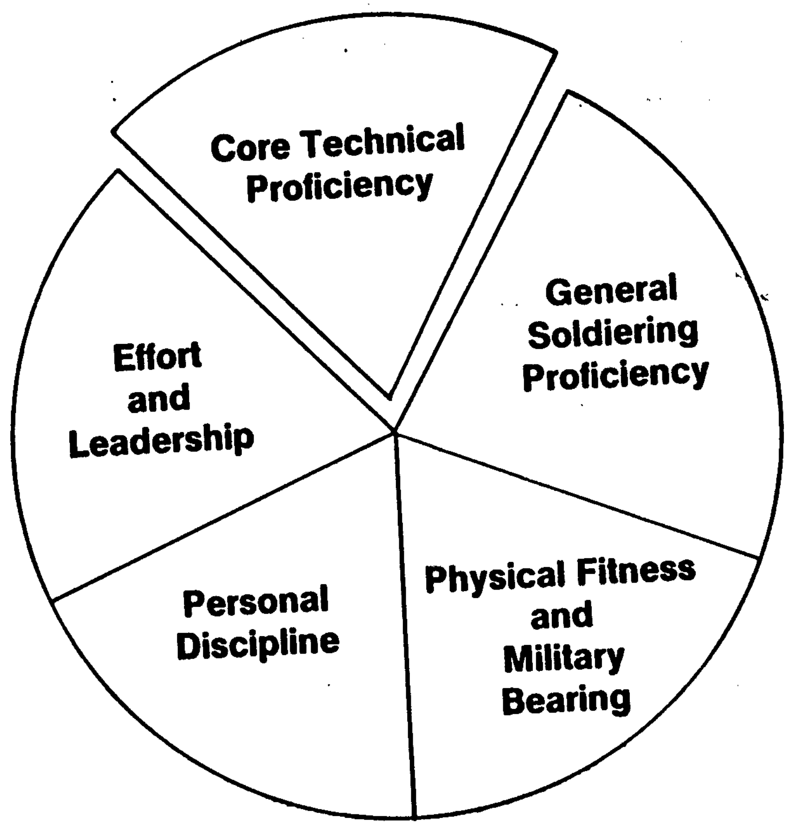
Ratings of Discipline & Self-Control
Avoiding Article 15
Being Promoted On-Time

Ratings of Physical Fitness
Military Appearance
Physical Readiness Scores

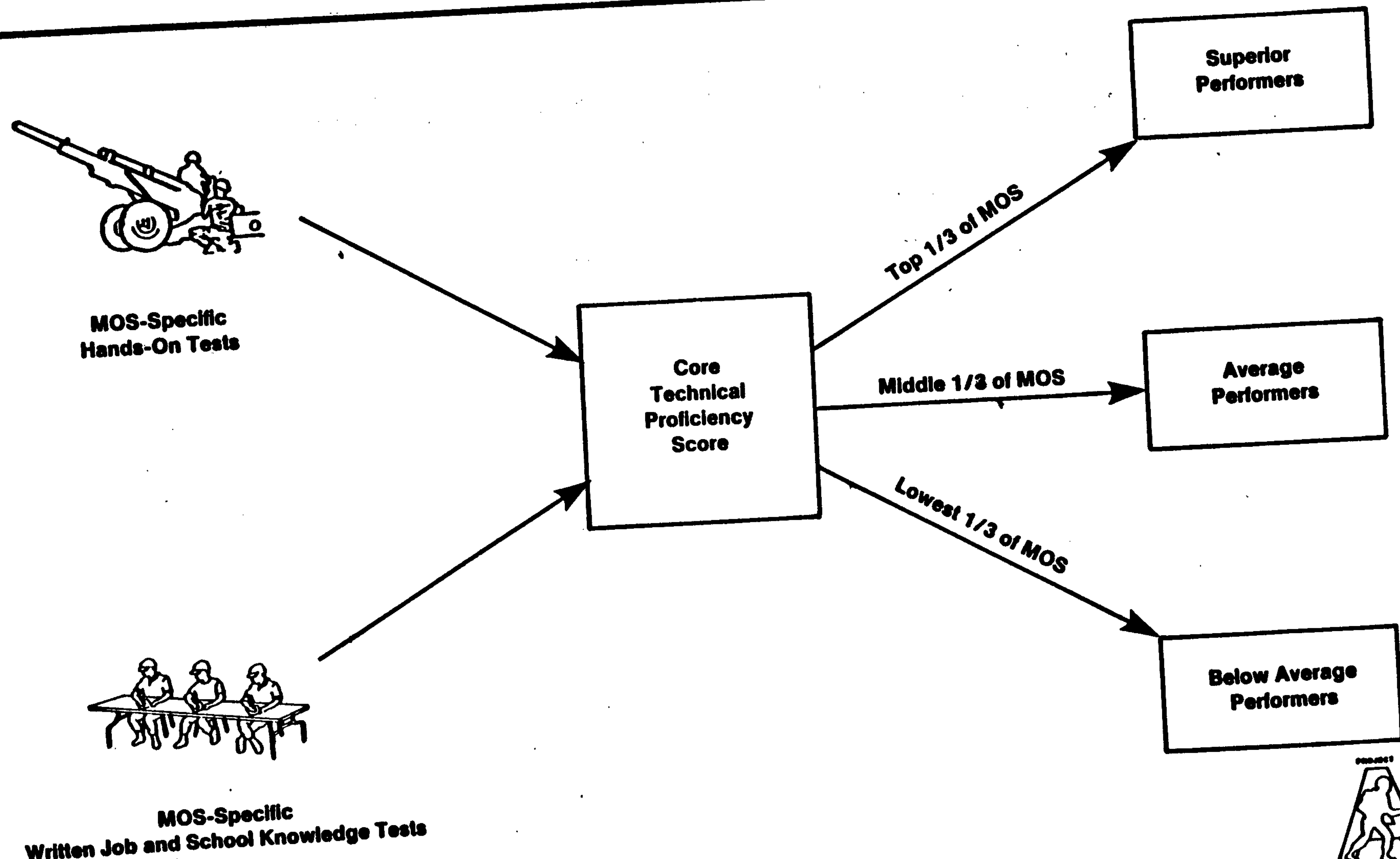
PROJECT A
JOB PERFORMANCE CONSTRUCTS



PROJECT A
JOB PERFORMANCE CONSTRUCTS



CORE TECHNICAL PROFICIENCY
IDENTIFYING SUPERIOR, AVERAGE, AND BELOW AVERAGE PERFORMERS



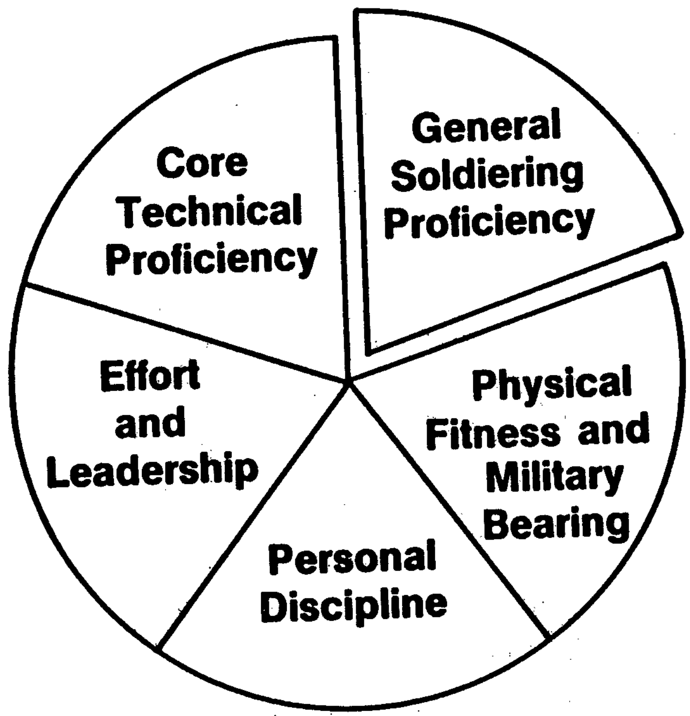
CORE TECHNICAL PROFICIENCY SUPERIOR, AVERAGE, AND BELOW AVERAGE

MOS 63B

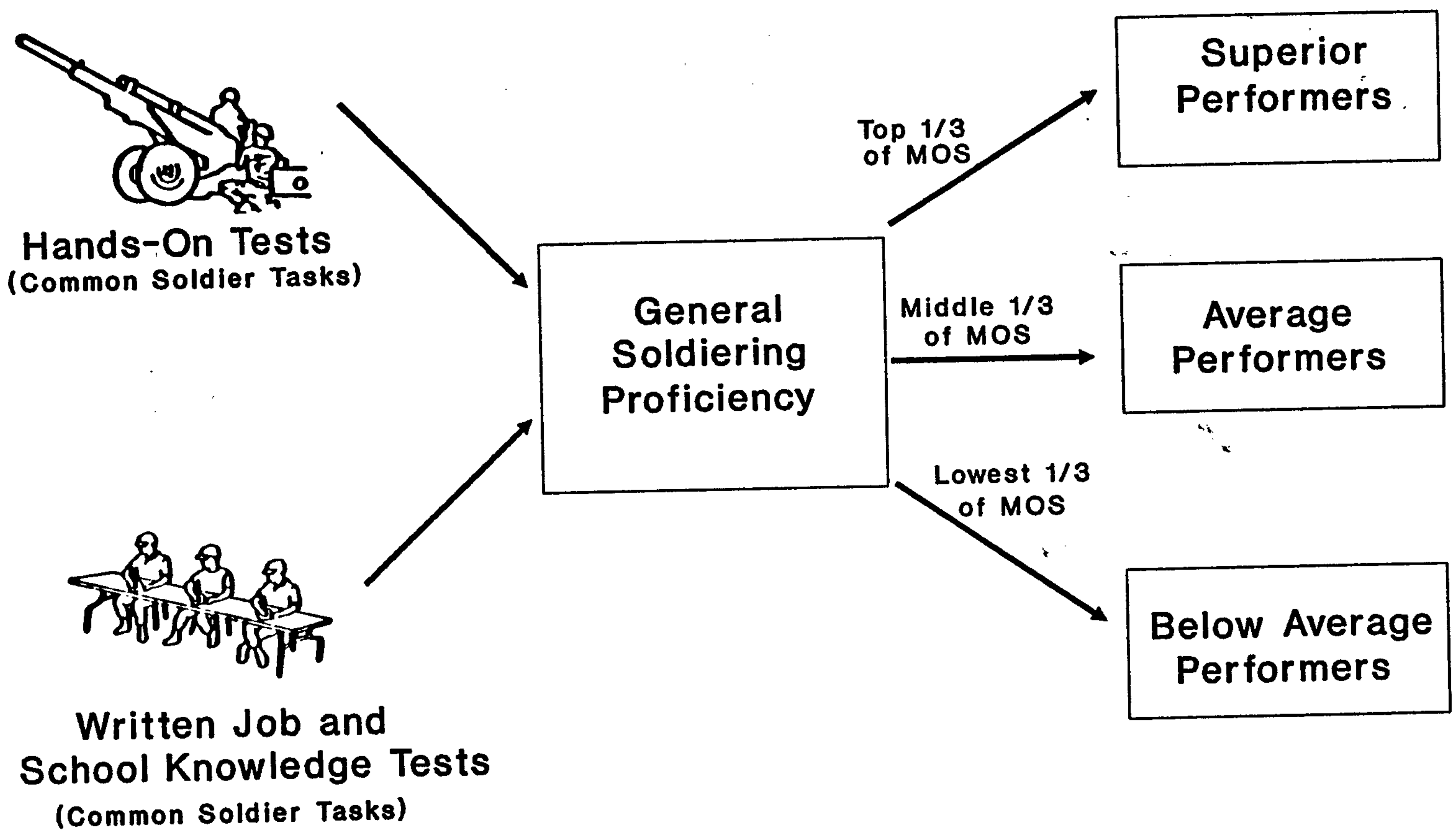
Task	Hands-On Test Mean Percent Go			Job Knowledge Test Mean Percent Correct		
	Superior	Average	Below Average	Superior	Average	Below Average
Adjust clutch pedal	90%	85%	76%	81%	68%	56%
Troubleshoot service brakes	89%	84%	73%	73%	67%	61%
Replace service brakes[1]				80%	70%	60%
Replace radiator[1]				77%	68%	53%

[1] These tasks were not tested hands-on

PROJECT A
JOB PERFORMANCE CONSTRUCTS



GENERAL SOLDIERING PROFICIENCY IDENTIFYING SUPERIOR, AVERAGE, AND BELOW AVERAGE PERFORMERS

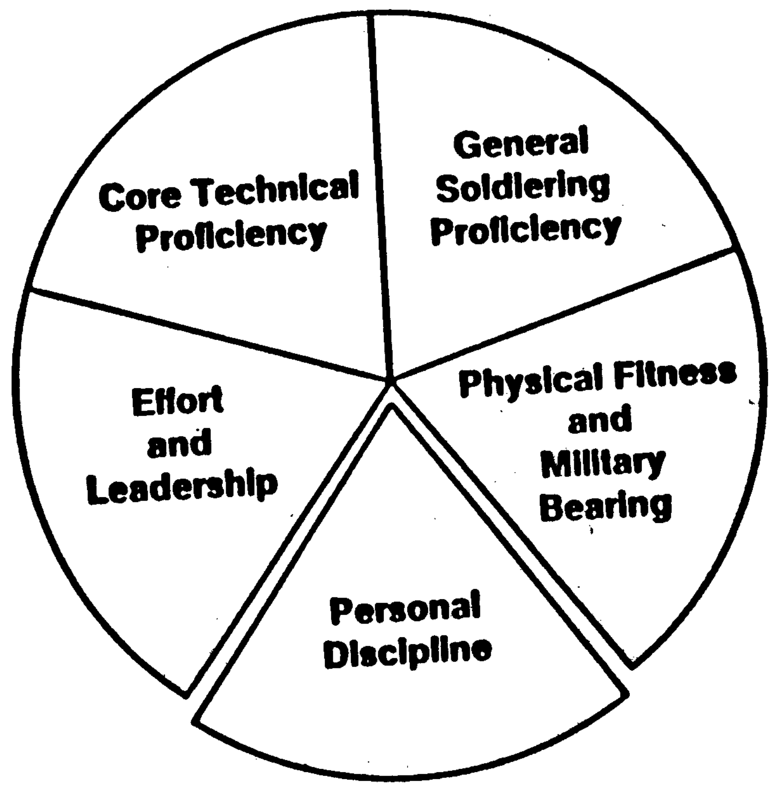


GENERAL SOLDIERING PROFICIENCY
SUPERIOR, AVERAGE, AND BELOW AVERAGE
MOS 63B

Task	Hands-On Test Mean Percent Go			Job Knowledge Test Mean Percent Correct		
	Superior	Average	Below Average	Superior	Average	Below Average
Determine magnetic azimuth using a compass	87%	79%	57%	55%	45%	34%
Put on field or pressure dressing	79%	72%	63%	63%	57%	52%
Perform annual PMCS[1]				79%	70%	61%
Tow disabled vehicle[1]				64%	54%	48%

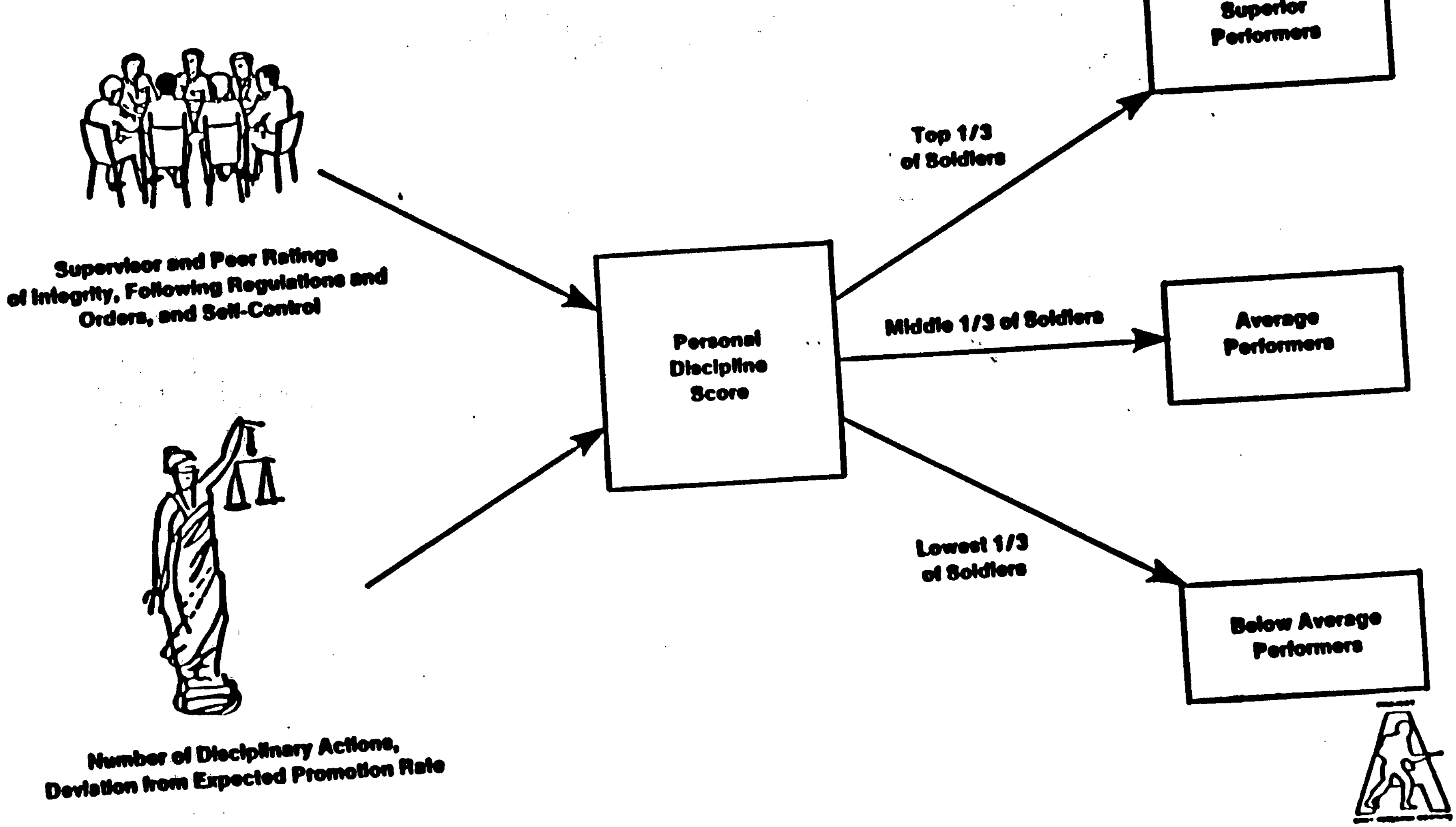
[1]These tasks were not tested hands-on

PROJECT A
JOB PERFORMANCE CONSTRUCTS



PERSONAL DISCIPLINE

IDENTIFYING SUPERIOR, AVERAGE, AND BELOW AVERAGE PERFORMERS



PERSONAL DISCIPLINE

NUMBER OF DISCIPLINARY ACTIONS FOR SUPERIOR, AVERAGE, AND BELOW AVERAGE SOLIERS

MOS 63B

Disciplinary Action	Mean Number of Actions		
	Superior	Average	Below Average
Articles 15	.02	.09	.58
Flag Actions	.01	.03	.28
TOTAL	.03	.12	.86

DESCRIPTION OF JOB PERFORMANCE LEVELS

SUMMARY

- Significant Differences between Below Average Soldiers and Superior and Average Soldiers
- Performance Differences Impact the Army's Readiness to Accomplish Its Mission
- Army Would Benefit by Recruiting More Superior Soldiers and Fewer Below Average Soldiers

PREDICTION OF JOB PERFORMANCE USING THE ASVAB

CORE TECHNICAL PROFICIENCY

MOS 63B

AFQT CATEGORY	Core Technical Proficiency Performance Level		
	Superior	Average	Below Average
I-IIIA	48%	26%	26%
IIIB	27%	40%	33%
IVA-IVB	12%	35%	53%

PREDICTION OF JOB PERFORMANCE USING THE ASVAB

GENERAL SOLDIERING PROFICIENCY

MOS 63B

AFQT CATEGORY	General Soldiering Proficiency Performance Level		
	Superior	Average	Below Average
I-IIIA	44%	33%	23%
IIIB	31%	36%	33%
IVA-IVB	11%	31%	58%

VALIDITY COEFFICIENTS OF PERFORMANCE WITH ASVAB AND ASVAB + NEW PREDICTORS

	MOS Technical	General Soldiering	Effort/ Leadership	Personal Discipline	Military Bearing and Physical Fitness
ASVAB Only	63	65	31	16	20
ASVAB + New Predictors	67	70	44	37	42
Gain	04	05	13	21	22

THE ASSESSMENT OF BACKGROUND AND LIFE EXPERIENCES (ABLE)

- Measures Temperament and Personal History
- Consists of 199 Items
- Requires 25-35 Minutes to Administer
- Yields Four Overall Scores
  - Achievement Orientation
  - Dependability
  - Physical Condition
  - Adjustment

PREDICTION OF JOB PERFORMANCE USING THE ABLE

EFFORT AND LEADERSHIP
MOS 63B

ABLE Achievement Orientation Score	Effort and Leadership Performance Level		
	Superior	Average	Below Average
Top 1/3	50%	31%	19%
Middle 1/3	27%	36%	37%
Lowest 1/3	23%	33%	44%

PREDICTION OF JOB PERFORMANCE USING THE ABLE

PERSONAL DISCIPLINE

MOS 63B

ABLE Dependability Score	Personal Discipline Performance Level		
	Superior	Average	Below Average
Top 1/3	45%	33%	22%
Middle 1/3	37%	30%	33%
Lowest 1/3	18%	36%	46%

PREDICTION OF JOB PERFORMANCE USING THE ABLE

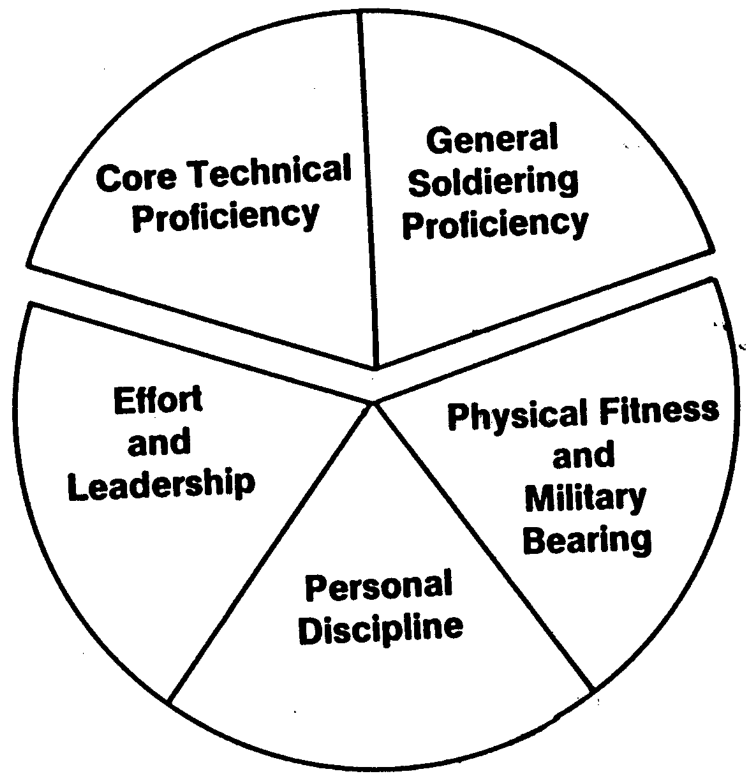
PHYSICAL FITNESS AND MILITARY BEARING

MOS 63B

ABLE Physical Fitness Score	Physical Fitness & Military Bearing Performance Level		
	Superior	Average	Below Average
Top 1/3	46%	32%	22%
Middle 1/3	33%	37%	30%
Lowest 1/3	21%	32%	47%

PREDICTION OF JOB PERFORMANCE

Excellent Prediction Using the ASVAB



Excellent Prediction Using the ABLE

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-21

Usefulness of Spatial and Psychomotor Testing for Predicting TOW and UCOFT Gunnery Performance

Jay M. Silva
Army Research Institute

December 1989

REVIEWED BY:

Amy C. Schwartz
Research Psychologist
Selection and Classification Technical Area

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification Technical Area

CLEARED BY:

for Zita M. Simutis, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Usefulness of Spatial and Psychomotor Tests for Predicting TOW and UCOFT Gunnery Performance

Jay M. Silva

General Introduction

This paper describes a predictor validation effort undertaken by the Army Research Institute with the assistance of the Centers and Schools of the U. S. Army Infantry at Fort Benning and U. S. Army Armor at Fort Knox. The goal was to examine whether new tests, developed under an on-going effort to measure psychomotor and spatial skills, would improve the selection of trainees into gunnery Military Occupational Specialties (MOS) beyond the contribution already made possible by the Armed Services Vocational Aptitude Battery (ASVAB).

The ASVAB is presently the major selection battery used by the Services. At the time of enlistment, Army trainees were required to pass cognitive ability screens based on a combination of ASVAB subtests. The 10 ASVAB subtests are:

Arithmetic Reasoning (AR)
Auto and Shop Information (AS)
Coding Speed (CS)
Electronics Information (EI)
General Science (GS)
Mechanical Comprehension (MC)
Mathematics Knowledge (MK)
Numerical Operations (NO)
Paragraph Comprehension (PC)
Word Knowledge (WK)

In practice, the last two subtests are not used individually, but rather are combined into a single Verbal (VE) score. Each subtest is standardized to the 1980 youth population (Department of Defense, 1982) and possesses a mean of 50 and a standard deviation of 10. The Armed Forces Qualification Test is a composite of three of these subtests (i.e., AR + MK + 2*VE) that is used to screen applicants into the Army. A second, more tailored composite of the subtests is used for selection into a specific MOS. In examining the contribution made by ASVAB, the individual subtests rather than the composites are used in order to maximize the contribution ASVAB can make.

The new spatial and psychomotor tests were developed under Project A. The project was a comprehensive long term effort to improve the Army's selection and classification system for enlisted personnel. Specific goals included developing improved performance measures and developing selection tests which tapped

a broader range of the attribute domain than already covered by ASVAB. The new predictors were developed through an iterative process which included empirical validation. The data collected in the validation research was used to sharpen the usefulness of the tests as well as our understanding of the attributes they measured.

The new predictors used in the present study can be divided into two groups: spatial and psychomotor. The two spatial tests, Mazes and Orientation, measured spatial visualization and spatial orientation, respectively. The two psychomotor tests, One- and Two-Hand Tracking, measured psychomotor precision and two-hand coordination, respectively. A short description of these tests along with the operational measure used for each is presented in Table 1. More detailed descriptions are available in Grafton, Czarnolewski, and Smith (1989).

Table 1
Description of Spatial and Psychomotor Tests

Test	Target Ability Domain	Measure
Spatial		
Mazes	Spatial Visualization (Scanning)	Number correct
Orientation	Spatial Orientation	Number correct
Psychomotor		
One-Hand Tracking	Psychomotor precision	Distance accuracy
Two-Hand Tracking	Two-hand coordination	Distance accuracy

TOW Gunners

Introduction

After enlisting as 11X, Basic Infantrymen, and completing Basic Training at Fort Benning, Georgia, they were classified into MOS 11H[1], Heavy Anti-Armor Weapon Infantryman. For brevity they are usually called TOW Gunners. Near the end of their courses in 1988, these students underwent simulator testing of their gunnery and other skills related to operating the weapon.

[1]A recruit enlists with a "Training MOS." On completing Basic and Advanced Individual Training, the recruit is officially awarded the MOS.

In order to qualify in the MOS, recruits were required to exceed a minimum gunnery score on the simulator test.

Prior to TOW (Tube-launched, Optically-tracked, Wire-guided) gunnery training, predictor data were gathered on new recruits at the Reception Battalion. After gunnery training, gunnery performance data were collected on high-fidelity gunnery simulators. This allowed the predictors to be validated in a predictive design with relevant, high quality data.

The TOW anti-tank gunnery simulator requires the gunner to track a moving target (i.e., a target mounted on a moving vehicle) through an infrared optical device. The target moves along a reasonably level paved road at a constant velocity. Since it does not move in a vertical direction (i.e., only to the degree that the terrain is not level), these data may not generalize to other more demanding engagements. The gunner aligns the crosshair by swiveling the weapon horizontally. The qualifying trials were scored in sets of 10 target engagements. Each engagement was scored for accuracy on a scale of 0 to 100, and the set score was simply the sum of the scores for the ten engagements. A score of at least 550 on a set was required to qualify. However, the student could not qualify until the third set (i.e., there were two practice sets). After attaining the qualifying score, the student's testing was complete. A student who did not attain 550 in eight sets did not qualify for the MOS.

Data are presented for 229 trainees who were tested in 1988 on the new predictors during their first three days of active duty. Testing consisted of two computerized and two pencil-and-paper tests (Table 2). Scores on these new tests were not used to select the students, who, nevertheless, had been told before the testing that the scores would go into their records.

Two types of gunnery performance were measured: 1) score on first qualifying set, and 2) pass or fail the first qualifying set. The first measure is a gunnery accuracy score with practice held constant across trainees. The second measure is an indicator of the amount of training needed to pass the course and relates to the cost of training.

Results and Discussion

Table 2 presents all validities and reliabilities for the ASVAB and spatial/psychomotor predictors. ASVAB subtests were related to score and pass/fail on the first qualifying set. Arithmetic Reasoning (AR), Auto and Shop Information (AS), General Science (GS), and Mechanical Comprehension (MC) all significantly ($p < .05$) predicted the score on the first qualifying set. Mechanical Comprehension had the largest

Table 2
TOW Data: Predictor-Criterion Relationships for First Qualifying Set

Predictor[1]	Criterion[2]: Pass/Fail at First Qualifying Set	Criterion[2]: Score at First Qualifying Set	Reliability[3]
AFQT	.14*	.16*	.92
AR	.17*	.17*	.88
AS	.10	.16*	.85
CS	.04	.08	.75
EI	.03	.10*	.72
GS	.12	.15*	.84
MC	.07	.21*	.78
MK	.08	.13	.86
NO	.04	.00	.68
VE	.08	.10	WK=.89 PC=.72
Spatial			
Mazes	-.05	-.03	.70
Orientation	.03	.13	.70
Psychomotor			
One-Hand Tracking	-.18*	-.28*	.74
Two-Hand Tracking	-.25*	-.33*	.85

Note. N=229 in the 1988 implementation.

* $p < .05$.

[1]Positive correlations for the spatial tests indicate that those who obtained more items correct were more likely to have passed and achieved a higher score at the First Qualifying Set. Negative correlations for the psychomotor tests indicate that those who had less distance from target were more likely to have passed and achieved a higher score at the First Qualifying Set.

[2]Estimated reliability for qualifying score (based on present data; correlation of First Qualifying Set with previous set) is .63. The point-biserial correlation between the two criterion measures was .73. Of the 229 students, 198 passed on the First Qualifying Set. Mean and standard deviation of score at First Qualifying Set are 636.28 and 112.59, respectively.

[3]None of the predictor reliabilities are based on present data. ASVAB reliabilities are alternate forms reliabilities for Form 11a correlated with Form 9a (Palmer et al., 1988). They are representative since most trainees were tested on ASVAB forms 11-13 and the reliabilities vary to only a small degree across forms. Since the reliability of VE was not available, reliabilities for the two components of VE (i.e., PC and WK) are reported in its place. Reliabilities for spatial and psychomotor tests are test-retest (unpublished IPR materials by J. Toquam, et al., PDRI, 1986).

correlation with this criterion. However, for the prediction of pass/fail performance on the first qualifying set, only Arithmetic Reasoning contributed. The Armed Forces Qualification Test (AFQT) significantly ($p < .05$) predicted performance on both of these criteria.

Turning to the new tests, only the psychomotor tests (i.e., both One- and Two-Hand Tracking) significantly ($p < .05$) predicted performance on both criteria. In the sample, Two-Hand Tracking fared somewhat better for both criteria, although the differences between the two tracking predictors were not significant ($p > .05$). The cause for this difference at the sample level may be that the Two-Hand Tracking test, in the order presently administered, has been shown to be more test-retest reliable than the One-Hand Tracking test (Toquam, Peterson, Rosse, Ashworth, Hanson, and Hallam, 1986).

A backward regression approach to building a model for all criteria involved the following: 1) backward regressing on only the 9 Armed Services Vocational Aptitude Battery (ASVAB) subtests, 2) forcing in the previously identified ASVAB subtests and backward regressing on only the new spatial/psychomotor subtests, 3) removing any ASVAB subtests not significant at the .05 level and recalculating the model.

For both criteria there was a significant ($p < .05$) amount of validity added by the new tests above ASVAB subtests. For details see Tables 3 and 4. For the score on the first qualifying set, Two-Hand Tracking and Mazes predicted an additional 9.6 percent of the criterion variance (i.e., over a 200 percent increase in explained criterion variance). Since Mazes did not, by itself, significantly predict score on the first qualifying set, its impact was felt because it suppressed variance in the other predictors. This is confirmed in Table 5 where it can be seen that Mazes significantly correlates with both MC and Two-Hand Tracking.

The criterion of pass/fail on the first qualifying trial was also significantly enhanced ($p < .05$) by Two-Hand Tracking. In terms of Somer's D_{yx} (i.e., an index of rank correlation between predicted probabilities and observed outcomes), adding Two-Hand Tracking to ASVAB raised the index from .292 to .486, for an increase of .194.

In summary, in predicting TOW gunnery performance indices, the usefulness of Two-Hand Tracking was consistently high. Mazes was the only other useful predictor but its impact was only as a suppressor.

Table 3
TOW Data: Predicting Score at First Qualifying Set

Model	Predictor	R	F	p
ASVAB Only		.214	10.95	.0011
	MC			
ASVAB Plus New Tests		.377	12.44	< .0001
	MC	.142	5.30	.0222
	Two-Hand Tracking	-.304	24.26	< .0001
	Mazes	-.123	4.01	.0463
R^2 Increment		.0963	25.17	< .01

Note. N=229. Backward stepwise regression was used. No adjustments or corrections were applied to the data.

Table 4
TOW Data: Predicting Pass/Fail at First Qualifying Set

Models	Concordance	Somer D_{yx}	X^2	p
ASVAB Only	.646	.292	6.95	.0084
AR				
ASVAB Plus New Tests	.743	.486	17.05	.0002
AR				
Two-Hand Tracking				
Increment	.097	.194	10.10	< .01

Note. N=229. Backward stepwise logistic regression was used. No adjustments or corrections were applied to the data.

Table 5
TOW Data: Selected Predictor Relationships

	MC	AR	Two-Hand Tracking
AR	.41*		
Two-Hand Tracking	-.27*	-.16*	
Mazes	.16*	.11	-.24*

Note. N=229. * $p < .05$.

M1 Armor Crewman

Introduction

This section describes the performance of a sample of entry level trainees at Fort Knox in MOS 19K, M1 Armor Crewman, on spatial/psychomotor tests and school gunnery exercises in 1988. Scores on the new spatial and psychomotor tests were not used to select the trainees, who, nevertheless, had been told before the testing that the scores would go in their records.

Near the end of their course, these students underwent simulator testing of their gunnery skills. For this purpose, the Unit Conduct of Fire Trainer (UCOFT) high-fidelity tank gunnery simulator was used. The measure of performance used in the analyses is a speed/accuracy composite constructed by subtracting the standardized opening time (i.e., time from when a target appeared until the first round was fired) from the standardized percent hits. This score was then transformed into a T-score with a mean of 50 and a standard deviation of 10.

Results and Discussion

Table 6 presents all zero-order validities and reliabilities for the ASVAB and spatial/psychomotor predictors. Many of the ASVAB subtests that correlated with the TOW performance indices also correlated with the UCOFT speed/accuracy performance composite. The two largest correlations were with Auto and Shop

Table 6
UCOFT Data: Predictor-Criterion Relationships for Speed/Accuracy

Predictor[1]	Speed/Accuracy[2]	Reliability[3]
AFQT	.23*	.92
AR	.21*	.88
AS	.33*	.85
CS	.06	.75
EI	.03	.72
GS	.24*	.84
MC	.42*	.78
MK	.18*	.86
NO	-.06	.68
VE	.23*	WK=.89
		PC=.72
<u>Spatial</u>		
Mazes	.35*	.70
Orientation	.35*	.70
<u>Psychomotor</u>		
One-Hand Tracking	-.40*	.74
Two-Hand Tracking	-.43*	.85

<u>Note</u>. N=244 in 1988 implementation.

* $p < .05$.

[1]Positive correlations for the spatial tests indicate that those who obtained more items correct were more likely to have achieved a higher speed/accuracy score. Negative correlations for the psychomotor tests indicate that those who took less time to respond or had less distance from the target were more likely to have achieved a higher speed/accuracy score.

[2]Mean and standard deviation of speed/accuracy criterion are 50.54 and 9.19, respectively.

[3]None of the predictor reliabilities are based on present data. ASVAB reliabilities are alternate forms reliabilities for Form 11a correlated with Form 9a (Palmer et al., 1988). They are representative since most trainees were tested on ASVAB forms 11-13 and the reliabilities vary to only a small degree across forms. Since the reliability of VE was not available, reliabilities for the two components of VE (i.e., PC and WK) are reported in its place. Reliabilities for spatial and psychomotor tests are test-retest (unpublished IPR materials by J. Toquam, et al., PDRI, 1986).

Information and Mechanical Comprehension (.33 and .42, respectively). All new predictors were significantly related to the speed/accuracy score on UCOFT. Mazes and Orientation were related to the criteria to the same degree (i.e., 0.35). One- and Two-Hand Tracking were also related to the criterion to about the same extent (i.e., -0.40 and -0.43, respectively; see Table 8).

The backward regression approach used to build TOW prediction models was also used for the UCOFT analysis. In addition to the ASVAB subtests, both Two-Hand Tracking and Mazes added validity to predicting the speed/accuracy score on the UCOFT. Although Orientation and Mazes had equal zero-order correlations with the speed/accuracy score, Mazes shared more variance with speed/accuracy once both the ASVAB subtests and Two-Hand Tracking were included (see Tables 7 and 8).

In summary, the usefulness of Two-Hand Tracking was again strongly supported. Mazes also added predictive power, and for this criterion it added it through a direct relationship rather than by removing non-criterion related variance from the other predictors.

General Discussion

The tracking tests appear extremely useful for predicting the gunnery criteria under examination (i.e., both TOW and UCOFT). Whether One- or Two-Hand Tracking is the better predictor is unclear. Obviously, Two-Hand Tracking is more strongly related. However, because the order of administration was held constant (i.e., One-Hand Tracking followed by Two-Hand Tracking), it is not clear whether Two-Hand Tracking is a stronger predictor due to higher reliability arising from practice on One-Hand Tracking. Strength is added to this argument from the fact that reliability was higher for Two- than One-Hand Tracking. Regardless, tracking ability information is very useful for predicting future gunnery performance.

The picture for the two spatial predictors is not as consistent. Mazes, for example, predicts first qualifying TOW score and UCOFT speed/accuracy score but does so through different mechanisms: Once through suppressor effects and once through direct prediction. In addition, it does not predict TOW pass/fail on first qualifying set. Orientation was not useful for predicting TOW performance scores, but was useful for predicting UCOFT speed/accuracy score. Even though, compared to Mazes, Orientation added less unique variance once ASVAB was used, it nevertheless points to a relationship that may be of use if the ASVAB's composition changes.

Table 7
UCOFT Data: Predicting Speed/Accuracy Score

Model	Predictor	R	F	p
ASVAB Only		.448	30.28	< .0001
	MC	.145	27.65	< .0001
	AS	.303	6.33	.0125
ASVAB Plus New Tests		.553	26.29	< .0001
	MC	.163	9.10	.0028
	AS	.130	5.83	.0165
	Two-Hand Tracking	-.224	17.31	< .0001
	Mazes	.161	8.89	.0032
R^2 Increment		.105	36.05	< .01

Note. N=244. Backward stepwise regression was used. No adjustments or corrections were applied to the data.

Table 8
UCOFT Data: Selected Predictor Relationships

	MC	AS	Two-Hand Tracking
AS	.48		
Two-Hand Tracking	-.41	-.22	
Mazes	.27	.18	-.36

Note. N=244. All correlations are significantly different from zero at .05 Type I error.

Regardless of what new tests enhance the prediction of gunnery criteria, the new tests definitely provide a means to improve the selection of TOW and UCOFT gunners. Although the skills needed to operate weapons may be common to several weapons systems (e.g., as tracking skills appears to be for these two systems), it is also likely that specific skills will be needed for some weapons and not others. For example, evidence exists that the present tests do not predict performance on simulators of the Chapparall and Stinger air defense weapons (Gast and Johnson, 1988).

When performance across the two weapons systems is predicted to a different extent by the same tests, it may be signaling that different skills are needed for operating the weapons. For example, the UCOFT gunner is more isolated from the outside-the-tank environment and thus may rely more on the spatial visualization and orientation skills measured by the Mazes and Orientation tests. In addition to the predictive power they provide, these specialized ability tests can also be used for improving classification efficiency. Examination of other gunnery MOS may suggest hypotheses as to which aspects of the different weapons systems demand which skills.

References

Department of Defense. (1982). Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery. Washington, D. C.: Office of the Assistant Secretary of Defense (Manpower, Reserve, Affairs, and Logistics).

Gast, I. F., & Johnson, D. M. (November, 1988). Evaluating psychomotor and spatial tests for selecting air defense gunners. Paper presented at the 30th Annual Conference of the Military Testing Association.

Grafton, F., Czarnolewski, M. Y., & Smith, E. P. (January, 1989). Relationship between Project A psychomotor and spatial tests and TOW2 gunnery performance: A preliminary investigation (Report ARI-WP-RS-89-1). Alexandria, VA: Army Research Institute.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D. (1988). Armed Forces Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (Report AFHRL-TP-87-48). San Antonio, TX: Air Force Human Resources Laboratory.

Toquam, J., Peterson, N., Rosse, R., Ashworth, S., Hanson, M. A., & Hallam, G. (March, 1986). Concurrent validity data analysis: Cognitive paper-and-pencil and computer-administered predictors (Trial Battery). Unpublished data analyses by Personnel Decisions Research Institute, Minneapolis, Minnesota.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-7

Validity of AFQT and Education Level for Predicting Hands-On Performance in 11H TOW Heavy Anti-Tank Missilemen

Jay M. Silva
Selection and Classification Technical Area

August 1989

REVIEWED BY:

Leonard A. White
Research Psychologist
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief, Selection and
Classification Technical Area

CLEARED BY:

Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Validity of AFQT and Education Level for Predicting Hands-On Performance in 11H TOW Heavy Anti-Tank Missilemen

Jay M. Silva

This report was prepared as a response to a technical request from the Deputy Chief of Staff for Resource Management in the Training and Doctrine Command. This office is conducting research (viz., Soldier Performance Research Project) aimed at examining the continued use of the Armed Services Vocational Aptitude Battery (ASVAB). Toward this end, the project is seeking to examine the validity of the ASVAB applied to a broader range of performance measures. Hands-on performance data recently collected at Fort Benning provides one opportunity to address this issue.

The purpose of this report is to examine the relationship between the Armed Forces Qualification Test (AFQT) and educational attainment (i.e., high-school graduate or not) with performance on hands-on tasks for TOW (Tube-launched, Optically-tracked, Wire-guided) Heavy Anti-Tank Missilemen. Although other measures may predict this performance as well as or better than AFQT and educational attainment, the focus here is solely on these two predictors.

The 342[1] cases for these analyses took the ASVAB as civilians before entering the Army. After enlisting as 11X, Basic Infantrymen, and completing Basic Training (BT) at Fort Benning, Georgia, they entered Advanced Individual Training (AIT) in MOS 11H[2], TOW Gunner. At the conclusion of 13 weeks of training, the recruits were tested on seven hands-on tasks directly related to their effectiveness as TOW Gunners. The score on each hands-on task is the number of tries necessary to perform sufficiently well to pass.

Only four of the eight hands-on tasks were retained for analysis since only those have historically shown variation in scores across recruits. The four hands-on tasks are:

---

[1]An additional 178 recruits were dropped from the analysis due to missing data on one or more of the following pieces of information: AFQT score, high school diploma status, hands-on scores.

[2]A recruit enlists with a "Training MOS." On completing Basic Combat Training and Advanced Individual Training, the recruit is officially awarded the MOS.

Task	Description
1	Assemble the TOW launcher
2	Conduct system check out procedures and pre-operational inspection of the TOW launcher and encased missile
3	Determine if a target can be engaged
4	Prepare an anti-armor range card for TOW

Means, standard deviations, and maximum number of attempts needed to pass each hands-on task are presented in Table 1. Task 4 took, on average, the most attempts to complete, and also showed the largest amount of variability. Task 3 took the lowest number of attempts to complete and displayed the smallest amount of variability.

Table 1

Means and Standard Deviations On Selected TOW Hands-On Tasks

Hands-On Tasks	Mean	SD	Maximum Attempts
Task 1	1.24	.51	3
Task 2	1.19	.48	3
Task 3	1.03	.17	2
Task 4	1.45	.63	3
Sum of 4 tasks	4.91	1.01	8

Note. N was 342 for each task.

The high school diploma status was coded dichotomously (0=no diploma; 1=diploma). AFQT score is a percentile score based on the new AFQT composite initiated in January 1989 (i.e., 2VE + AR + MK). Table 2 lists the zero-order correlations of the two predictors and total hands-on performance. Only AFQT was related to total attempts to pass hands-on tasks.

Table 2

Means and Standard Deviations of Predictors and Zero-Order Correlation Matrix

	Mean	SD	Correlation: Total attempts to pass hands-on tasks	Correlation: AFQT
High school diploma status	.84	.37	-.103	-.070
AFQT	52.96	20.90	-.231*	

Note. N was 342 for each statistic. An asterisk indicates statistical significance at $p < .05$.

Regressing total number of attempts to pass the four hands-on tasks on high school diploma status and AFQT score yielded the results in Table 3.

Table 3

Regression of Total Attempts to Pass Hands-On Tasks On High School Diploma Status and AFQT Score

Predictor	Standardized Beta	SE of Estimate	F	p
High school diploma status	-.088	.053	2.75	.0980
AFQT	-.225	.053	18.18	< .0001

Note. N was 342 for each task. Negative signs indicate that trainees with higher predictor scores needed fewer attempts to pass hands-on tasks.

High school diploma status was not statistically related to total hands-on performance if a Type I error probability of .05 is used as the cutoff for significance. AFQT score, however, is related to this performance. In applied terms, this means that for every 1 standard deviation change in mean AFQT score, it is expected that the mean number of total trials to complete the hands-on tasks is reduced by .225 standard deviation units. For example, a one standard deviation increase in mean AFQT of 11H

trainees, can be expected to lead to a mean total trials of 4.68 versus the existing 4.91 now needed.

Conclusion

AFQT moderately predicts TOW Gunner total hands-on performance. By itself, high school diploma status has a much lower relationship with total hands-on performance, which strictly speaking is not significantly different from zero. Finally, adding high school diploma status to AFQT provides non-significant improvement.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER

WP-RS-89-4

Relationship of AFQT Category to TOW Gunnery Performance

Jay M. Silva and Clinton B. Walker
Selection and Classification Technical Area

May 1989

Approved for Distribution to Sponsor or Proponent

REVIEWED BY:

Elizabeth Brady
Research Psychologist, Selection and
Classification Technical Area

APPROVED BY:

Frances C. Grafton
Acting Chief, Selection and
Classification Technical Area

CLEARED BY:

Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

**U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600**

Relationship of AFQT Category to TOW Gunnery Performance

Jay M. Silva and Clinton B. Walker

This report describes the relationship between Armed Forces Qualification Test (AFQT) scores and performance on the M70 TOW anti-tank missile simulator for Infantry trainees. The goal is to examine whether the quality of soldiers, in terms of their AFQT scores, is reflected in their performance in a combat MOS. Although other measures may predict TOW gunnery performance as well or even better than AFQT, the focus here will be only on AFQT.

The 1,713[1] cases for these analyses took ASVAB as civilians before entering the Army. After enlisting as 11X, Basic Infantrymen, and completing Basic Training at Fort Benning, Georgia, they were classified into advanced training in MOS 11H[2], TOW Gunner. Near the end of their courses in 1988, these students underwent simulator testing of their gunnery skills. In order to qualify in the MOS, recruits were required to achieve a minimum passing score on the simulator test.

The M70 TOW anti-tank gunnery simulator requires the gunner to track the missile onto the target. The gunner aligns the crosshair by swiveling the weapon horizontally and turning two hand knobs for elevation. The qualifying trials on the M70 are scored in sets of 10 target engagements, a set being called an event. Each engagement is scored for accuracy on a scale of 0 to 100, the event score being simply the sum of the scores for the ten engagements. A score of at least 550 on an event is required to qualify, Event 3 being the first on which qualification is awarded. After attaining the qualifying score, the student's testing is complete. A student who does not attain 550 in eight events does not qualify for the MOS.

The relationship of AFQT to performance is revealed in analyses of three types of gunnery scores by AFQT Category: qualifying score, score on Event 3, and number of events needed to reach the qualifying level. These measures represent two important aspects of performance, respectively: accuracy, with practice held constant across trainees; and amount of practice

---

[1]An additional 47 recruits did not ever achieve a qualifying score. The analysis of these recruits is presented separately in the Appendix. The reason they were not included in the main analysis was that it was not clear why they had not qualified (e.g., 96 percent of these recruits did not use all 8 events to attempt to qualify). Eighteen cases did not have AFQT category available.

[2]A recruit enlists with a "Training MOS." On completing basic and advanced individual training, the recruit is officially awarded the MOS.

needed to attain qualification. The latter measure is an indicator of training costs.

Table 1 shows by AFQT category the means and standard deviations of the qualifying and Event 3 scores. Statistical comparison of these mean scores across AFQT categories finds no significant differences among categories 1, 2, and 3A. However, when categories 1, 2, and 3A were collapsed and compared to a collapsed composite of 3B and 4A, a significant difference emerged ($p < .01$). For the qualifying score, the estimated unweighted mean difference between these two groupings of AFQT categories was 15 points less for the lower AFQT categories. For Event 3 score the difference was 36 points less for the lower AFQT categories.

The fact that qualifying scores, compared to Event 3 scores, exhibited less of a mean difference between AFQT categories was a result of the tradeoff that exists between gunnery performance and practice. Those in the lower AFQT categories used practice to elevate their scores.

Table 2 shows the percentage of trainees who qualified in the minimum number of events (i.e., Event 3) and the percentage who took an unusually large number of events to qualify. Trainees in category 4A were less likely to qualify at Event 3 and were more likely to need at least 5 events to qualify ($p < .05$).

These data support the conclusion that general ability, in terms of AFQT, is related, although to a small degree, to the critical combat skill of gunnery. Trainees in AFQT categories at or below 3B achieve a lower accuracy score at the time of qualification; and a larger portion of trainees in category 4A take more time to qualify.

Table 1

TOW Gunnery Scores by AFQT Category

AFQT Category	N	Qualifying Score Mean	Qualifying Score SD	Event 3 Score Mean	Event 3 Score SD
1	73	654	69	610	123
2	612	653	74	627	111
3A	391	646	69	613	113
3B	531	637	67	597	119
4A	106	635	67	564	155

Table 2

TOW Gunnery Qualification Event by AFQT Category

AFQT Category	N	Percent Not Qualifying On Event 3	Percent Needing More Than 4 Events to Qualify
1	73	19	7
2	612	15	4
3A	391	16	8
3B	531	21	7
4A	106	29	12

Appendix

Forty seven of the recruits apparently did not qualify for the MOS based on obtaining a score of 550 or better on Event 3 or later. It is not clear why these recruits did not qualify. The following tabulations indicate why the picture is not clear: 1) 45 recruits (95 percent) did not use all 8 events to attempt to qualify, 2) 8 recruits (17 percent) did not attempt even one event, and 3) of the ones who attempted at least one event, only 25 (64 percent) even attempted to qualify on more than 3 events.

A descriptive analysis was conducted on these 47 cases in order to explore the relationship that may exist between AFQT classification and TOW gunnery performance. Caution should be excercised in interpreting the results due to extremely small sample size within AFQT category. The mean highest score achieved and mean number of events completed in each AFQT category is presented in Table A1.

Table A1

Mean Highest Score And Number of Completed Events by AFQT Category

AFQT Category	Number Failing to Qualify	Mean Highest Score Achieved	Mean Events Completed
1	0	-	-
2	9	488	4
3A	12	468	3.6
3B	22	493	4.8
4A	4	522	5.3

Note. AFQT category was not available for 18 of the recruits.

Selection and Classification Technical Area

Working Paper RS-WP-88-07

Data Analysis to Support the Committee on the Performance of Military Personnel

ELIZABETH P. SMITH & ANDREA E. BIRNBAUM

August 1988

Approved for Distribution to Sponsor or Proponent.

REVIEWED BY:

MICHAEL G. RUMSEY
Research Psychologist

APPROVED BY:

LAWRENCE M. HANSER
Chief, Selection and
Classification Technical Area

Manpower and Personnel Research Laboratory

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

Encl 3

DATA ANALYSIS TO SUPPORT THE COMMITTEE ON THE PERFORMANCE OF MILITARY PERSONNEL

Elizabeth P. Smith & Andrea E. Birnbaum

The analyses discussed here with the attached computer output were generated at the request of the National Academy of Science's Committee on the Performance of Military Personnel for the Joint-Service Job Performance Measurement Project. An effort was made to conform to the specifics of the request as closely as possible. However, some changes were necessary due to constraints imposed by the data and by available resources.

Data Base

The data were taken from the Project A Concurrent Validation Sample[1] for those Military Occupational Specialties (MOS) for whom Hands-on Test Scores were available. These MOS and the number of soldiers in each were:

Infantry (11B)	489
Cannon Crewmember (13B)	451
Tank Crewmember (19E)	390
Radio Operator (31C)	288
Vehicle Mechanic (63B)	476
Motor Transport Operator (64C)	499
Administrative Specialist (71L)	424
Medical Specialist (91A)	391
Military Police (95B)	590
Total	3998

The variables used in the analyses were the ten Armed Services Vocational Aptitude Battery (ASVAB) subtest scores, Armed Forces Qualification Test (AFQT) score, the MOS-specific Aptitude Area Composite (AA) (i.e. Service Composite) score and the following Project A performance total test scores:

- Hands-on (all tasks)
- Hands-on (MOS-specific tasks)
- School Knowledge (K3)
- Job Knowledge (K5)
- MOS-specific Behaviorally-Anchored Rating Scale (BARS)
- MOS-specific Task Ratings

Reliability values (Recommendation 8) for the Job Knowledge Task Tests (K5), the Hands-on Task Tests (all tasks) and the MOS-specific Behaviorally Anchored Rating Scales (BARS), can be found on page 5-7 of the Project A 1986 Annual Report (Campbell, 1986). Reliability values for the School Knowledge Tests (K3) are shown in Table 1 (data provided by S. R. Schultz, D. B. Kuhn, and C. B. Walker, personal communication, December, 1987).

The usual standardized scores for the ASVAB, AFQT, and AA scores were used. The Project A test scores were converted to standard T-scores

(mean = 50, standard deviation = 10) by MOS to enable comparisons across the same metric. Not all data were available for all subjects and no special attempts were made to impute test scores for subjects on these measures.

Analyses

The first set of analyses calculated the frequency distributions, range, mean, median, standard deviation, skewness, kurtosis, and additional information for each of the 18 variables by MOS. The output for this set is labelled "Descriptive Statistics by MOS." These results respond to Recommendation 6 of the "Proposals for Common Data Analysis in the Joint-Service Job Performance Measurement Project."

The second set of analyses determined the correlations and covariances among the 18 variables for each MOS. The output is labelled "Correlations for ASVAB, AFQT, AA, and Performance" and "Covariances for ASVAB, AFQT, AA, and Performance." These results respond to Recommendation 9a.

In response to Recommendation 9b, first we plotted each variable with all others by MOS. This output is labelled "Plot of --- * ---" where "---" represents a variable. Second, we performed polynomial regressions by MOS. That is, we regressed each variable on every other variable using the following model: $Y=X+X^2+X^3+X^4$. The difference between the R^2 obtained for this model and the squared product-moment correlation reflects departure from linearity. We did not compute these departures from linearity, but have provided the necessary information to do so. These printouts are labelled "Check for Linearity: --- and ---" indicating the variables involved.

References

Campbell, J. P. (Ed.) (1986). Improving the selection, classification, and utilization of Army enlisted personnel: Annual Report, 1986 Fiscal Year. (ARI Tech. Report No. 813101). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Footnote

[1]An earlier report on Project A Concurrent Validation data [Arabian, J. M., Wise, L. L., & Young, W. (1988). Army research to link standards for enlistment to on-the-job performance: 6th Annual Report to Congress. (ARI Working Paper RS-WP-88-2). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences] presented results based on analyses which excluded subjects with missing scores on certain criterion variables. Such an exclusionary rule was required for that report because of the manner in which criterion construct scores were generated. The analyses examined here did not require the exclusion of such subjects. Accordingly, these results are based on different sample sizes from those reported in Arabian, et al. (1988).

Table 1

Results From JRKTs Administered During Concurrent Validation

MOS	Number of Subjects	Number of Items	Mean Percent Correct	SD	Range	Alpha	Mean Number Correct
Batch A							
11B	491	150	62.1	11.4	93	.93	93.2
13B-SP	371	136	57.0	10.1	72	.88	77.5
13B-T	86	156	54.8	11.3	96	.89	85.5
19E	394	162	68.7	10.8	110	.93	111.3
31C	289	175	60.3	11.4	110	.93	105.6
63B	478	171	61.0	11.3	94	.93	104.4
64C	507	128	63.1	10.6	78	.90	80.8
71L	427	105	59.2	10.7	65	.86	62.2
91A	392	175	60.3	10.3	103	.92	105.6
95B	597	139	59.4	9.6	79	.87	82.5
Batch Z							
12B	544	176	58.9	11.5	114	.93	103.7
16S	338	149	70.5	9.0	82	.89	105.0
27E	123	175	56.8	13.6	102	.94	99.4
51B	69	163	62.8	13.6	108	.94	102.4
54E	340	145	68.7	10.3	85	.90	99.5
55B	203	181	60.1	13.3	122	.95	108.9
67N	238	175	66.3	8.3	81	.89	116.0
76W	339	175	49.8	9.8	95	.90	87.3
76Y	444	175	61.9	10.9	119	.93	108.4
94B	368	175	57.3	10.3	99	.92	100.2

PB 9114

SELECTION AND CLASSIFICATION TECHNICAL AREA

SC

WORKING PAPER WP-~~RS~~-89-6

ARI's Nominations of New Tests for the Technical Advisory Selection Panel (TASP)

Clinton B. Walker
Selection and Classification Technical Area

July 1989

REVIEWED BY:

Frances C. Grafton
Statistician
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief, Selection and
Classification Technical Area

CLEARED BY:

for Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

WP SC 89-6

ARI's Nominations of New Tests
For the Technical Advisory Selection Panel (TASP)

Clinton B. Walker, Editor

General statement

This report forwards to TASP ARI's nominations of new tests for use in the ECAT cost/benefit project. ARI has in its databases the information to address almost all of the issues in TASP's criteria for future tests. However, a number of useful analyses will not have been completed before the TASP proposal on new tests has to go forward.

So far under Project A, analyses of Army's new ability tests have been done at the level of composites of predictors, not at the level of the individual test. The data from Project A are archived and await analysis under a follow-on contract called Manning the Career Force. Bids have been received on it, and a startup by October 1989 is expected. By the time that the ECAT cost/benefit project is ready for evaluation, the relevant Army analyses of Project A data will have been completed. These analyses will cover fairness, alternative scoring methods (e.g., trading off speed and accuracy to defeat testwise response strategies), and other items in the TASP evaluation criteria.

Here is a brief description of the data that ARI has archived from Project A and that will support the relevant analyses. Most of Army's performance measures for JPM have been collected. Project A conducted its Concurrent Validation (CV) in 1985, when over 9,000 first tour soldiers in 19 MOS took all of the new predictors. These MOS were picked with the advice of a panel of General Officer advisors to be representative and critical. The result was a set of MOS that account for 45% of Army's annual accessions.

Nine of the MOS in the CV, these supplying over 5,000 cases, were given hands-on tests of MOS-specific and general soldiering tasks. Validities have been computed so far for constructs or composites like "Spatial," "Computerized Tests," and "Hands on Performace." Incremental validities relative to pre-enlistment ASVAB were computed for these predictor composites. In those analyses, the ASVAB score was a linear composite of four factors: Technical, Quantitative, Verbal and Speed.

The next phase of Project A was the Longitudinal Validation (LV), in which 50,000 new recruits took the new predictors and then had their success in Service tested at two subsequent points: end of entry-level job training (knowledge tests for all 21 MOS; n >40,000) and first tour on the job (hands-on and knowledge tests for 9 MOS, total n about 6,000, and knowledge tests for 11 other MOS, total n about 4,000).

In addition to data from Project A, ARI has results from smaller projects that have been carried out with trainees in tank and anti-tank gunnery. For these smaller projects, some analyses at the level of the individual test have been done and are summarized in this report. In regression analyses of these gunnery data, some tests give inconsistent results. This variation may be due to changes in the set of variables in the analyses, changes in the particular test score used (e.g., speed or accuracy), changes in the conditions of predictor or performance testing, or to variations in samples of examinees. So these analyses are exploratory.

On a scale of task criticality, hitting targets is near the top; so the value of these smaller projects for TASP is their measures of gunnery performance from high-fidelity simulators of weapon systems. Data from a new implementation of three spatial tests will be available by the end of June. These tests are being piloted as predictors of success in land navigation, a task that is important in many Army jobs.

Incremental validity for selection

Although TASP is looking at new tests mainly for classification, the minutes to January's meeting note that gains in overall validity should be assessed as well. It is conceivable that even small gains in pan-occupational validity could enlarge the expected utility of new tests enough to win continued support for them. In this regard, the CV has found overall gains in predicting the two constructs of maximal performance, MOS-specific Technical Proficiency and General Soldiering Proficiency. Both of those outcomes are composites of training knowledge, on-the-job knowledge, and on-the-job hands-on skills.

In the CV, the average R for all ASVAB subtests against MOS-specific Technical Proficiency was .63. Adding the new computerized tests to the ASVAB subtests gains a point; adding the six new spatial tests gains two points. For the outcome of General Soldiering Proficiency, the R for all of ASVAB was .65. Spatial tests added three points to that, computerized tests two points. Against the hands-on component of job-performance, ASVAB had a validity of .47 compared with .49 each for ASVAB plus the spatial and computerized predictors. Other combinations of the predictors need to be examined. For all analyses reported in this paragraph, a multivariate correction for restriction of range was applied to ASVAB and a correction for shrinkage used.

The criterion problem

Predictor tests are not good candidates for incrementing validity by themselves; their potential benefits are with respect to specific performance measures or classes of such. In order for the new tests in the ECAT cost/benefit project to receive a fair tryout, they need to be validated against appropriate criteria. This concern is prompted in part by the move to use

end of training grades as criteria for a major part of the ECAT cost/benefit project. Those grades come mostly from knowledge tests, which we already know are well predicted by ASVAB. A plan to rely heavily on end of training grades is a plan not to look very hard for incremental validity.

In the same vein, ARI has found strong relations between new Project A tests and performance on two weapon systems (Grafton, et al., 1989; Graham, 1988; Smith & Graham, 1987; Smith and M. Walker, 1988), but poor validities for some of the same predictors against performance on a different weapon system (Gast & Johnson, 1988). Another difference is between validities for predicting performance of Tank Crewmen: when the criterion is a composite of hands-on tasks, the entire new Project A battery (i.e., spatial, computerized, and non-cognitive) correlates .43, but when the criterion is tank gunnery alone, the correlation of two spatial and two computerized predictors is .54. So if the fate of the future tests rests on their showing good incremental validities in the ECAT cost/benefit project, the choice of performance measures is critical.

A relatively low cost tactic for getting objective data on MOS-specific hands-on tasks is to include occupations that routinely qualify new trainees and/or requalify incumbents on high-fidelity simulators. In two such occupations, ARI has found sizable increments in validity against simulated gunnery performance. The cost of putting part-time contract employees in the field at several posts to collect routinely generated performance measures might be small compared to the benefit and to the alternative tactics for getting criterion data.

One uncertainty that TASP will have to live with is the life expectancy of high-fidelity simulators. Budget cuts could curtail routine performance testing on these devices at the end of training.

A related issue

How does one interpret the data from high-fidelity simulators? There is some concern that validity for predicting a specific, highly critical measure, such as accuracy in gunnery, may not represent validity against the totality of performance on the job. That concern is legitimate, but a reverse concern is equally important: performance on a set of tasks that has been composed for randomness, convenience, or criticality from (e.g.) a gunner's job will not tell how well the gunner can shoot if shooting is not one of the tasks. Since classification is our main concern, we may be justified in overlooking randomness, breadth, etc., in our task sampling. Looking at a few MOS-unique, highly critical tasks may be sufficient for some MOS. ASVAB can take care of the gunner's knowledge and general skill; the new tests can tell how well the gunner can hit the target.

Utility of gains in classification

An item that may need early discussion is how to make a case for the utility of a test's gains in classification. Classification may be justified by psychometric evidence or by evidence that the net utility of recruiting, training, and success/failure in some jobs is greater than in others. Meanwhile, each Service puts some constraints on classification in order to distribute able persons across jobs. These constraints are meant to insure that all occupations can provide themselves able leaders from within. But constraints on distribution limit the impact of a test on classification, and thus the possible gains in utility. The challenge then is to find out whether there will be a net gain in utility through improved classification in spite of these constraints. Information on Services' systems for distributing quality personnel may be needed.

Organization of Army entry-level training

Recruits in many occupations enter the Army at one training post, take Basic Combat Training there, and then go to a different post for training in their MOS. If the ECAT validation is designed to administer predictors at RTCs/Reception Battalions and to have the end of training as one source of performance measures, then the two kinds of data may have to come from different posts for some Army occupations. In the case of 11H TOW Gunners and 19K Tank Crewmen, students take their basic and technical training at the same post. But for the other MOS that Army is offering, 13F Fire Support Specialist, students entering technical training at Fort Sill come from several basic training posts.

The more new tests the better

Since the time for predictor testing in the ECAT cost/benefit project is limited, TASP may not be able to include all of the new tests that show promise. One block of testing time that bears examining is the one for CAT-ASVAB. ARI believes that there will be a much better payoff from the cost/benefit project by devoting all the testing time to new tests than by dropping new tests so that CAT-ASVAB can be administered concurrently. Here's why:

A close facsimile of concurrent CAT-ASVAB scores will be available for each examinee: pre-enlistment, pencil/paper ASVAB scores. For any new test that is excluded from the ECAT battery there will be no joint-Service facsimile. Moreover, CAT-ASVAB will have another opportunity to interact with new tests in the large scale data collection that would be needed to bring new tests into the ASVAB. But any new test that is dropped now will not have another chance to compete for a place in Joint Service pre-enlistment testing for a very long time. The statistical interplay of the new tests (e.g., redundancies, suppressor

effects) is as important as their individual validities, so it is essential to have as many of them as possible included now.

<u>Irrelevant response strategies</u>

A number of tests are proposed in this paper because they will help identify examinees' special abilities for tasks in combat MOS. Those MOS are hard to fill owing partly to their large size and partly to the lack of equivalent civilian jobs. Applicants could well try to steer themselves away from combat MOS and toward MOS with civilian counterparts by doing poorly on, e.g., psychomotor tests. In this time of the all volunteer forces, we usually do not expect examinees to "fake bad," but such faking needs to be added to the research agenda for the new tests.

TESTS OF SPATIAL ABILITIES (pp. 8 - 37)

Assembling Objects
Mazes
Orientation
Map
Reasoning

This section of the report describes five of the six spatial tests from the Project A battery. Validities of the individual tests against the Project A performance constructs have not been computed yet. In the table below are validities and incremental validities for a unit weighted composite of all six spatial tests against a composite of MOS-specific hands-on tasks in Project A.

Project A Concurrent Validation
Validities Against Hands-on Performance in Nine Major MOS
For ASVAB and Spatial Predictor Composites

	Occupations									
Predictor Composites	11B	13B	19E	31C	63B	64C	71L	91A	95B	ALL
ASVAB (4 factors, unit wtd)	.44	.22	.40	.65	.30	.54	.54	.58	.59	.47
All Spatial (unit wtd)	.39	.24	.40	.51	.26	.53	.53	.55	.57	.44
Spatial + ASVAB	.45	.24	.43	.64	.33	.57	.58	.60	.61	.49

Source: McHenry, Szenas, & Wise (1987, unpublished)

Names of the occupations are as follows:

11B	Infantryman	64C	Motor transport operator
13B	Cannon crewman	71L	Administrative specialist
19E	Armor crewman	91A	Medical specialist
31C	Radio teletype operator	95B	Military police
63B	Light-wheel vehicle mechanic		

It now appears that the performance measures for validating new tests in ECAT will be solely end-of-training knowledge tests, at least in some occupations. An indication of the power of Army's new spatial tests to predict scores on knowledge tests (qua pencil/paper multiple choice tests) is given in the next table. There, a composite of end-of-training knowledge tests and on-the-job knowledge tests in the Project A CV is the criterion.

Unlike end-of-training tests that are a part of training courses, the Project A knowledge tests were designed to have ample variances. Thus, the correlations in the following table are probably an optimistic estimate of the validities to be expected in the ECAT validation.

Project A Concurrent Validation
Validities Against Knowledge Tests in Nine Major MOS
For ASVAB and Spatial Predictor Composites

Predictor Composites	11B	13B	19E	31C	63B	64C	71L	91A	95B	ALL
ASVAB (4 factors, unit wtd)	.77	.61	.76	.77	.80	.74	.73	.81	.78	.75
All Spatial (unit wtd)	.70	.62	.72	.66	.70	.67	.68	.70	.69	.68
Spatial + ASVAB	.80	.66	.79	.77	.81	.75	.76	.82	.80	.77

Outline of case for Assembling Objects to be in ECAT Validation

Description: Assembling Objects is a somewhat speeded 36 item test from the spatial battery of Project A. The item "stem" is a figure showing a set of separate parts. Examinees select from multiple choices the one that is a possible assemblage of the parts. The parts assemble into a regular geometrical outer border or into a tinker-toy like design. Directions and sample items are given below.

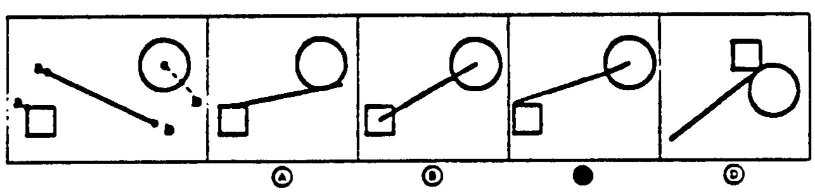
ASSEMBLING OBJECTS TEST

In this test you are to figure out how an object will look when its parts are put together correctly.

For each problem, the first picture shows the parts that need to be put together. The next four pictures, labeled A, B, C, and D, show four different ways the parts could be put together. Only one of them is correct.

There are two types of problems in the test. In the first type, the parts are labeled with letters, and by matching the letters, you can see where the parts should touch when the object is put together correctly. Look at Example 1.

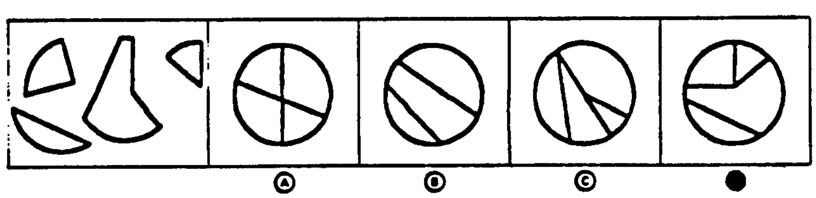
Example 1:



First, look at the square with the corner marked "a". Imagine what the object would look like if you attached that corner to the end of the line marked "a", and then attached the "b" end of the line to the middle of the circle, where the "b" is. Of the four answer choices, labeled A, B, C, and D, only C shows the object put together correctly, so circle C has been filled in.

For the second type of problem in this test, the parts to be put together are not labeled. In these problems, you just fit the parts together like the pieces of a puzzle. Only one of the four answer choices is correct. Look at Example 2 below.

Example 2:



When you piece together the parts in the first picture, only answer choice D can be correct. D is the only answer choice that contains all the right shapes or pieces of the puzzle.

To mark your answer, fill in the lettered circle below the correct answer choice, as shown in the two examples above.

1. **Theoretical development of construct and measure(s)**

A. **Definition of construct (process[es])**
Domain: Spatial abilities
Sub-domain: Spatial visualization (Rotation)

The ability to restructure pieces of a two-dimensional display by mentally manipulating them and judging the results. Rotation, here two dimensional, is the more specific ability to identify a two-dimensional figure in different orientations in the plane of the page.

How is this construct related to other constructs/ factors? In Alderton's discussion of relations among spatial tests, he places the Minnesota Paper Form Board, a test whose items resemble the bounded figure items of Assembling Objects, at the "complex" and "power" ends of the dimensions of complexity and speed vs. power. For correlational data, see below 3.A. Construct Validity.

B. **Taxonomy/types/categories of items**
Two kinds of items are on this test. In the first, the parts of the object are all closed figures and the object that they form is a recognizable geometrical shape (e.g., circle or polygon). In the second type, the parts include lines and angles along with small closed figures. Response choices then look like small tinker-toy assemblages, or abstract mobiles. In the stem, the parts are labelled with letters indicating where they connect with the correspondingly lettered components of other parts. The letters define and limit correct arrangements of the parts together.

How the test is scored/scaled: Total number correct

How are item type and difficulty jointly sampled: Eighteen of each type are on the test. Examination of design and statistical parameters for the items is needed to finish answering this question.

C. **Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):** Likely SES related differences in exposure to abstract diagrammatic stimuli.

D. **Basis for expecting incremental validity in specific jobs:** "Predictive validity estimates provided by expert raters [Wing, Peterson, and Hoffman, 1984] suggest that measures of the visualization construct would be effective predictors of success in MOS that involve mechanical operations (e.g., inspect and troubleshoot mechanical [or] electrical systems), construction... and drawing or using maps. (p. 3-5)" (Peterson, 1987)

Alderton (1989) reports that the Minnesota Paper Form Board, which has items quite like the second item type on AO, has produced significant validities against grades in courses for following jobs: aircraft inspector, air traffic controller, navigator, sewing machine operator, pharmaceutical packer, auto mechanic, carpenter, gunboat and motor-launch operator, signal[person], and plumber. These jobs either resemble or appear to have important elements in common with enlisted Army jobs.

2. **Precision of measurement**

A. **Test's consistency across**

Different forms: One form exists now.

Testing occasions: Retest reliability was .70 for > 468 Army incumbents. In 67 Marines, the value was .51 (source: 4/88 DAC briefing).

Levels of total test score (score conditional SEM): To be determined under the contract that follows up Project A.

B. **For power or speeded test: internal consistency:** Alpha of .90 in 9,000+ incumbents; .91 Split half, with Spearman-Brown correction). In the Marine sample (n = 67), .89.

C. **Variation across subgroups in score-conditional SEM:** To be determined under the contract that follows up Project A.

D. **If score is model-based, what's the model-based precision:** N.A.

E. **Subgroup differences in item functioning:** Fairness analyses on this test will be done under the follow-on contract to Project A. Group means and standard deviations in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	6,007	24.75	6.04
Blacks	2,551	19.79	6.99
Hispanics	334	23.87	6.19
Other	355	23.97	6.27
Females	887	23.44	6.33
Males	8,360	23.31	6.73

3. **Validity**

A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** In the field test of the Project A predictor battery, Assembling

Objects and a marker test for the construct, FIT Assembly, had a correlation of .64 in 56 examinees.

Alderton (1989) supports his position that the Minnesota Paper Form Board is a power test of high complexity by showing that its intercorraltions with other tests of varying speededness and complexity are as they should be. The second item type in AO resembles those on the MPFB.

Assembling Objects is the test that loads most heavily on the spatial factor among Project A tests and ASVAB subtests (p. 81). In addition, it has no high loadings on other factors in the analyses of pre-enlistment ASVAB and the Project A battery. Among ASVAB subtests, its highest correlations are with MC, AR, and MK (p. 86).

B. **Uniqueness re ASVAB. How assessed?** Using internal consistency (.91) or retest reliability (.70) minus the squared multiple R of .26 with all ASVAB subtests, uniqueness is .65 or .44, respectively. Paul Mayberry's April 88 briefing to DAC reported somewhat lower reliabilities and uniqueness based on 67 cases.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, Project A tests during their first three days of Service, and tests of anti-tank gunnery skills about twelve weeks later. The criterion measure was an accuracy score on the fifth event (i.e., set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance.

AO correlated .082 with Event 5 scores ($p = .1305$) and was not significantly associated with unique criterion variance in analyses that include or excluded ASVAB subtests and ASVAB-related variance ($p = .082$ in the latter case). The multiple R^2 for all ASVAB subtests was .041, for all Project A was .166, and for both sets together .177. Project A tests accounted for .126 of the residual variance. In the stepwise regression ($p = .15$ for entry) for predicting residual variance, AO is the

fourth of five Project A tests to enter (p. 100).

The Marines have found in JPM work to date that Assembling Objects adds .061 in explained variance on top of four ASVAB factors, MOS, time in service, and time in service squared against a knowledge test, in a sample of >1,100 Marines (unpublished data reported to JPM 5/89). In addition, it adds .032 to the explained variance in predicting hands-on performance. Among the predictors in the Marines' battery were a couple of psychomotor measures, "spatial," Project A's Reasoning test, and the Armed Services Applicant Profile.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?** To be determined under the follow on contract to Project A.

E. **Job families and patterns of validities for them. Similarity of job families in other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:** For incremental validities of a composite of Project A spatial tests against major outcomes in nine MOS, see above pp. 6 - 7.

4. Equating

A. **Problems in developing equivalent forms (e.g., How large is the potential item pool?) Would pretesting and vetting of new items be needed? Could item devel be automated?** The potential item pool is unlimited. Over the short run, additional items could be created by rotating whole items or squares of information within items. Pretesting of new items would be advisable. Item development probably could be automated, but not cheaply.

B. **Types of samples needed/acceptable for initial equating? testing conditions needed?** Sufficiency in cells of minority, female, and low aptitude examinees would drive sample sizes

C. **Subgroup dependence of equatings:** To be determined at the point where we get into equating.

D. **Practicality of pencil-and-paper versions:** Assembling objects started as a pencil-and-paper test.

5. **Feasibility for operational S/C:**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees:** There is a fixed time for answering the items of 16 minutes. Time for giving directions is about 5 minutes. Dave Alderton (personal communication, 1989) reports the following time data

from administering a computerized AO to 1,305 Navy recruits. Navy's software permitted examinees to skip questions. Skipped items were automatically presented again at the end of the test, if time allowed.

Component\	mean	s.d.	maximum time
Instructions	4.75 min	1.46	14.85
Items	10.66	1.99	12.22
Total	15.41	2.82	26.85

B. **Susceptibility to:**

Cheating, compromise: Estimated to be low: verbal descriptions of the items probably would not be easy to memorize.

Irrelevant response strategies: On their face, the items seem to invite using a verbal response strategy. The findings that Assembling Objects anchors the spatial factor in the Concurrent Validation and that it adds a useful amount of variance on top of ASVAB in Marine research show that verbal strategies do not automatically wash out the test's unique contribution.

Coaching: Unknown; guessed to be low

Practice: An earlier version of Assembling Objects which had four more items showed a gain .08 s.d. in 502 incumbents (p. 97).

C. **Floor/ceiling effects? Performance of low/high AFQTs on the test?** On retesting, the mean score of 23.96 was 1.8 s.d. below a perfect score. Performance of different AFQT levels is TBD in the follow on to Project A.

D. **Examinees' reactions: motivation, enjoyment, boredom:** Nothing noteworthy.

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** None. This test is up and running on the HP. NPRDC has two forms of the executable software.

F. **Need for special eqpt with special maintenance?** None

Outline of case for the Map Test to be in the ECAT Validation

Description: Map is a somewhat speeded 20 item, multiple choice test from the spatial battery of Project A. On it, the examinee uses a schematic map and explicit compass directions to infer implicit ones. Every item has the same eight compass directions as the response choices. Directions and a sample item are given below.

MAP TEST

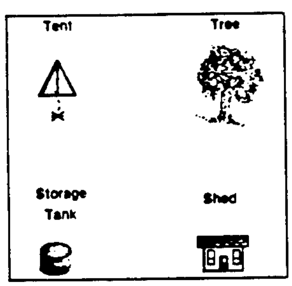
This test measures how well you can find your way around a landscape, using a map and your knowledge of compass directions.

There are two parts in this test. For each part, there is a map showing an aerial view of the ground. The map shows places like a barracks, a campsite, a forest, a lake, etc. For each test question, you will be given some indication as to which direction is north, west, east or south. For example, you may be told, "The barracks are due **SOUTH** of the storage tanks." Now you know which direction is South and you can use that to figure out all the other directions. Remember, as you face North, East is always to your right.

Next, you will be told that you need to get from one place to another, for example, from the forest to the campsite. You must decide which direction you would have to travel to get there, and then fill in the circle next to the correct answer.

Look at Example 1 below. In Example 1, you are told that the shed is directly North of the tree and that you are at the storage tank. If the shed is North of the tree, then which direction do you go to get from the storage tank to the tent? You must travel South (S), so circle ⑤ next to this answer, has been filled in. Please study this example for a minute to see for yourself why South is the correct answer, and then work Example 2 on your own.

Example Map:



1. **The shed is due north of the tree. You are at the storage tank. Which direction must you travel to reach the tent?**

   ① N ② NE ③ E ④ SE ● S ⑥ SW ⑦ W ⑧ NW

2. **The tent is due west of the storage tank. You are at the storage tank. Which direction must you travel to reach the tree?**

   ① N ② NE ③ E ④ SE ⑤ S ⑥ SW ⑦ W ⑧ NW

The answer for Example 2 is ⑧ , for NW or Northwest. If the tent is west of the storage tank, then the shed is north of the storage tank, and the tree is northwest of the storage tank.

Are there any questions about how to complete this test? Your score will be the total number of correct answers. Work as quickly and as accurately as you can. You will have 12 minutes to work on this test.

1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es])

Domains:

Spatial ability - Orientation. This test measures the ability to maintain one's bearings with respect to the compass despite frequent changes in orientation.

Deductive reasoning. Given an explicit direction from one site on the map to another in the item stem, the orientation of the compass can be deduced and used to deduce the direction from any third site to any fourth.

How is this construct related to other constructs/factors
This construct was not included in Project A's expert judgment task (Wing, Peterson, & Hoffman, 1984), so its location in the hierarchical map of the predictor space (Peterson, 1987, p. 118) did not emerge from that work. See below Construct Validity, 3.A.

B. Taxonomy/types/categories of items

Items vary in the extent to which non-cardinal directions are involved in the given direction and the question. Except for the names of the sites (the sites are clearly depicted and labelled on the map) and the specified directions, the syntax of the item stems is identical, with one minor exception.

How the test is scored/scaled: Total correct

How are item type and difficulty jointly sampled
The degree of difference between the given direction and North affects item difficulty, as well as the degree of difference between the given and deduced directions. Distribution of items over those conditions awaits an examination of the item specifications.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers)

Offensiveness is not an issue, as the items are almost content-free. Experience with maps, compass directions, and abstract diagrams may favor higher SES examinees. Rural and urban examinees may have experiences that could help, but those experiences are likely to be different.

D. Basis for expecting incremental validity in specific jobs:

The Map Test resembles the common soldiering tasks (i.e., tasks that are a part of every Army enlisted job) of identifying terrain features on a map and using them to navigate over land. In addition such tasks are a daily part of the job in Infantry, Air Defense, Armor, Field Artillery, Air Traffic Control, and Intelligence,

where it is essential to the work to be able to maintain directional orientation and to determine locations by using landmarks and known directions.

2. Precision of measurement

A. Test's consistency across

Different forms: Map has one form at present

Testing occasions: .78 in 499 incumbents

Levels of total test score (score conditional SEM): To be determined under the contract following up Project A

B. For power or speeded test: internal consistency

Alpha = .89; split half (Spearman-Brown) = .90

C. Variation across subgroups in score-conditional SEM:

To be computed under the contract that follows up Project A.

D. If score is model-based, what's the model-based precision?

N. A.

E. Subgroup differences in item functioning:

Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the number correct measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	6,007	9.32	5.53
Blacks	2,550	4.16	3.47
Hispanics	335	6.32	4.96
Other	355	7.01	5.40
Females	885	6.32	4.90
Males	8,362	7.84	5.56

3. Validity

A. Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:

Spatial and general abilities: Page 81 shows Map loading almost as heavily on the ASVAB factor as on the Spatial one. Its highest correlates among ASVAB subtests are MC, AR, MK, and GS (p. 86).

B. Uniqueness re ASVAB. How assessed?

The squared multiple R with all ASVAB subtests is .44. In terms of internal consistency (.90) and retest reliability (.78), uniqueness is .46 and .34, in order.

C. <u>Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?</u> In analyses of TOW gunnery data, all ASVAB subtests, and all Project A computerized and spatial tests, Map did not account for unique variance nor enter significantly into a stepwise regression (p. 100). By the end of June, we expect to have data evaluating the validity of Map in predicting performance in land navigation in a sample of 300 canditates for Special Forces.

D. <u>Subgroup differences in incremental validity or in std err of prediction of the criterion?</u>
To be calculated from the Project A data in the follow-on contract.

E. <u>Job families and patterns of validities for them. Similarity of job families in other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:</u> The combat MOS in which map reading and land navigation are critical tasks are large fill, costly-to-fill MOS. Intelligence MOS are small, costly to fill, and important. For incremental validity of a composite of Project A spatial tests against major outcomes in nine MOS, see above pp. 6 and 7.

4. <u>Equating</u>

A. <u>Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?)</u> The pool of potential items is limitless: by varying the map alone, the item stems alone, or both, we could create a vast pool of items. Item development from an algorithm on software seems practical. Pretesting of items probably would be needed to assure that item statistics were as expected.

B. <u>Types of samples needed/acceptable for initial equating? Testing conditions needed?</u> Status as applicant or recruit is probably not as important as the range of abilities in the sample. It should be as broad as that in the applicant population on which the tests will be used.

C. <u>Subgroup dependence of equatings:</u> To be determined once equating research is undertaken.

D. <u>Practicality of pencil-and-paper versions:</u> The present version is p/p and has been administered to 60,000+ examinees in Project A without difficulty.

5. Feasibility for operational S/C:

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees: Twelve minutes are allowed for working the items. Time for instructions is about 5 min.

B. Susceptibility to

Cheating, compromise: Given a large item pool, these should be low.

Irrelevant response strategies: Some examinees use a strategy of physically rotating the map. Whether that helps is not known. The strategy of using the given direction to determine North, then holding a finger or thumb pointed toward that North, seemed to help one ARI researcher score 2 s.d. above the mean on Map for first tour soldiers, but that strategy still requires the examinee to know what to do with the knowledge of where North is.

Coaching: Unknown

Practice: On retesting, 502 incumbents increased their mean score by .08 s.d. (p. 97).

C. Floor/ceiling effects. Performance of low/high AFQTs: With 8 response choices and 20 items, chance performance is 2.5 correct. That is less than one s.d. below the mean of 7.67 (s.d. 5.51, n >9,300). Separate analyses have not been done yet with AFQT levels.

D. Examinees' reactions: motivation, enjoyment, boredom: Nothing exceptional

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:
This test is up on the HP. The graphics are very clear, but the map is a bit crowded. No other hardware is required. The regular number or letter keys may suffice to enter the eight response choices.

F. Need for special eqpt with special maintenance?
None.

Draft of the case for using the Mazes test in the ECAT Validation

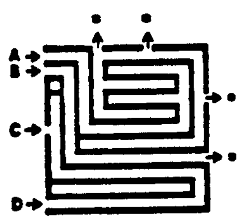
Description: Mazes is a somewhat speeded 24-item test from the spatial battery of Project A where the examinee selects the one entrance to a rectangular maze that leads to an exit. For directions and a sample item, see below.

MAZE TEST

This test measures how well you can find your way through a maze. Each maze in the test has four entrances, labeled A, B, C, and D. The exits are marked with an arrow pointing out, and a star (*). Only one of the four entrances leads to an exit out of the maze. You must figure out which entrance (A, B, C, or D) leads to a way out. Mark your answer by filling in the correct lettered circle below the maze.

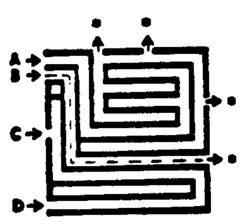
Look at the example below. Which entrance leads to a way out of the maze? Fill in the correct lettered circle below the maze.

EXAMPLE:



Ⓐ Ⓑ Ⓒ Ⓓ

The correct answer to this example is B. B is the only entrance that leads to a way out of the maze. The dotted line in the picture below shows how the entrance marked B is the only one that leads to an exit from the maze. To mark your answer, you should have filled in the circle B located below the maze, as shown. Please do not make any other pencil marks on the test.



Ⓐ ● Ⓒ Ⓓ

Are there any questions about how to complete this test? Your score will be the total number of correct answers. Work as quickly and as accurately as you can. You will have 5½ minutes to work on this test.

DO NOT BEGIN UNTIL TOLD TO DO SO.

1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es])

Domain: Primarily spatial ability

Spatial Sub-domain: It is not clear which of Alderton's big four spatial factors best fit(s) the abilities measured by this test. Perceptual Speed is the closest, Mazes involving speeded visual search. But the stimuli are complex in some ways, and the abilities to scan a field, stay on a path, and to remember which paths have been tried already (or to distinguish signal from noise and suppress the noise) are involved, too. TASP, at its April meeting, put Mazes under the non-committal heading "Spatial Relations." PDRI, which developed the Project A predictor battery, put Mazes under Spatial Visualization, which TASP explicitly did not.

How is this construct related to other constructs/factors
See below 3.A. Construct Validity

B. Taxonomy/types/categories of items
Item difficulty varies with lengths of the paths, in terms of number of turns and number of blind alleys.

How the test is scored/scaled
One score is used: total number correct.

How are item type and difficulty jointly sampled
To be determined in the upcoming follow on to Project A.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers)
Being content-free, the items are not offensive. There may be SES-related differences in exposure to posers, puzzles, or maze-like video games.

D. Basis for expecting incremental validity in specific jobs: ASVAB does not currently measure the spatial abilities which rational analysis suggests are needed in many Army jobs. Project A experts (Wing, Peterson, and Hoffman, 1984) judged that measures of the visualization and scanning constructs would have good validities against a variety of inspecting, troubleshooting, and repairing tasks (median .31), as well as using maps and controlling air traffic (~.38)

2. Precision of measurement

A. Test's consistency across

Different forms: Mazes has one form at present

Testing occasions: Retest reliability = .70 in 499 incumbents

Levels of total test score (score conditional SEM): To be determined under the contract that follows up Project A.

B. For power or speeded test, internal consistency: Odd-even reliability: .96 (N = 9,344 incumbents); Alpha: .89 (N = > 9,331 incumbents).

C. Variation across subgroups in score-conditional SEM: To be determined under the contract that follows up Project A.

D. If score is model-based, what's the model-based precision
N. A.

E. Subgroup differences in item functioning

Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the number correct by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	6,007	17.64	4.31
Blacks	2,551	13.36	4.50
Hispanics	335	16.92	4.03
Other	354	16.97	4.81
Females	885	15.04	4.81
Males	8,362	16.55	4.74

3. Validity

A. Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB

Spatial factor/construct
On page 81, the six spatial tests in the Concurrent Validation formed the second factor in a principal factor analysis of ASVAB subtests and new Project A tests.

Relation of Mazes to G
Page 81 shows Mazes to have the lowest loading of the spatial tests on the factor consisting of the eight unspeeded tests of ASVAB; viz., .12.

Spatial speed:

Correlated with ASVAB subtests, Mazes' highest correlations are with MC, MK, and AR, while the lowest correlations are with WK and the two speeded subtests, PC and NO (p. 86). Factor analyses of the six spatial tests from Project A gave a two-factor solution (p. 88) where Mazes clusters with the more speeded spatial tests, while a three-factor solution (p. 89) groups it with Object Rotation, the simplest, most speeded of the tests.

Other: On p. 89, Mazes is seen to have a strong secondary loading on a factor with Assembling Objects and Reasoning.

Psychomotor abilities

In factor analyses involving spatial and computerized tests from Project A (p. 81 including ASVAB and p. 83 without ASVAB), Mazes has the highest loading of the spatial tests on the psychomotor factor. In a separate sample of 95 Armor officers who took the spatial and psychomotor tests (p. 84), Mazes loaded more strongly with tracking than with the other five seemingly spatial tests. These latter two results are consistent with the observation that many examinees physically trace paths through the mazes with their fingers.

B. Uniqueness re ASVAB. How assessed?

The squared multiple R of all ASVAB subtests, given before enlistment, with Mazes was .22 for 7,329 first tour soldiers. Uniqueness was estimated to be .74 and .48 for internal consistency and retest, respectively.

C. Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated? Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all

Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance.

Mazes correlated .108 with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. In various GLM analyses with all other Project A tests entered (p. 100), Mazes did not account for significant unique variance nor significantly enter into the stepwise regression.

In 1988, a battery of four Project A tests -- One- and Two-hand Tracking, Mazes, and Orientation -- was installed at two training bases for a pilot test on students in the 11H anti-tank and 19K tank crew courses. In samples of these students (pp. 101-2), the validity of an optimally weighted combination of all ASVAB subtests against scores on TOW and anti-tank gunnery simulators was .266 and .480, respectively. Mazes had a validity of .020 (n.s.) and .308. In both of the 1988 samples, Mazes did account for significant unique variance and enter significantly into stepwise regressions on the residualized performance variance. In these regressions, it added .017 and .024 in variance accounted for.

D. <u>Subgroup differences in incremental validity or in std err of prediction of the criterion?</u>

For the sample in 2.A above, there were 88 blacks and 444 whites. Score-conditional SEP are as follows:

SD units of # correct on Mazes:	-2	-1	0	+1	+2
SEP W (n = 444)	1.5	0.9	0.6	0.8	1.4
SEP B (n = 88)	3.3	2.2	2.2	3.3	4.9

E. <u>Job families and patterns of validities for them</u>

For incremental validities of a composite of Project A spatial tests against major outcomes in nine MOS, see above pp. 6 - 7.

The value of new predictors for classification will be more obvious at the level of relevant tasks, not at the level of whole jobs.

<u>Similarity of job families in other services</u>

Most Army jobs have counterparts in the Marines. Tasks that involve inspecting, troubleshooting, and repairing electronics and mechanical systems have counterparts in all other Services.

Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants
Army's combat MOS (Infantry, Armor, Field Artillery, Combat Engineer) require large numbers and are the most critical (on a scale that equates criticality with closeness to combat), but are hard to fill. For at least one MOS in Infantry, Armor, and Artillery (11H, 19K, and 13F), success in "gunnery" is objectively measured near the end of entry-level training. For one other Infantry MOS, 11C Mortar Crewman, firing is supposed to be tested objectively twice a year in the active forces. Requalification on weapons systems is carried out periodically, either in live-fire or on simulators, in the active forces.

4. Equating

A. Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?)
Equivalent forms should be easy to create. First, "new" items could be made by rotating the present items 90, 180, or 270 degrees. Then, an algorithm for generating items could be written that varies the number of turns, distance between turns, etc. New items should have some pilot testing to make sure that their statistics are as expected.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed?
Filling cells for minorities, women, and low AFQTs will influence sample size. Conditions suitable for any pencil/paper or computerized testing should suffice.

C. Subgroup dependence of equatings
To be determined when equating research is undertaken.

D. Practicality of pencil-and-paper versions
High. Mazes originated as a pencil-and-paper test.

5. Feasibility for operational S/C

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees
Mazes is a speeded test with 5.5 minutes allowed to work the scored items. Time for directions is about 5 minutes.

B. Susceptibility to:

Practice: In 502 incumbents, mean scores on Mazes rose .20 s.d., from 16.54 to 17.46.

Irrelevant response strategies/coaching:
One tactic that works for some examinees, viz., work back from the exits instead of forward from the entrances, seems not to change the ability being measured. For items where more than one exit can be reached by the correct entrance, that strategy could give an advantage.

Cheating/compromise: None known

C. Floor/ceiling effects? Performance of low/high AFQTs on the test? Examinees averaged 1.6 s.d. below the ceiling in the Concurrent Validation (n > 9,331 incumbents). In research on tank and anti-tank gunnery (pp. 100 - 103), scores on Mazes had enough variance to correlate significantly with gunnery performance.

D. Examinees' reactions: motivation, enjoyment, boredom: Nothing remarkable

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required: Uses a great deal of space. Needs no additional hardware

F. Need for special eqpt with special maintenance: None.

Outline of case for Orientation to be in the ECAT Validation

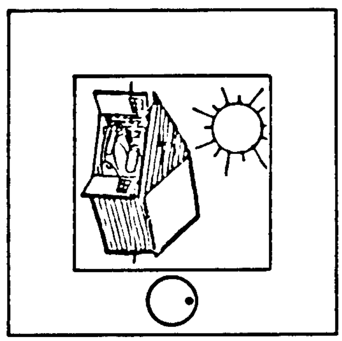
Description: Orientation is a somewhat speeded 24-item test from the Project A spatial battery where examinees select an answer depicting what a figure will look like after it is rotated in the plane of the page. Directions and sample items are given below and on the next page.

ORIENTATION TEST

This is a test of your ability to imagine how an object will look when it has been turned or rotated.

Please look at the sample problem below. It contains a picture with a frame around it. The bottom of the frame has a circle with a dot inside, carved into the frame.

EXAMPLE 1:



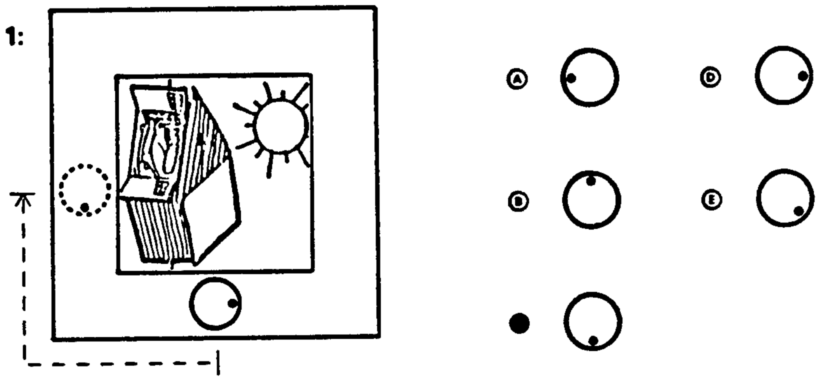
If you examine the frame and the picture inside it, you will notice that the bottom of the frame, marked by the circle with a dot, does not match up with the bottom of the picture. Imagine that only the frame can be turned, and the picture inside can not be moved. Then, to match up the bottom of the frame with the bottom of the picture, you would turn the frame until the circle with a dot is located at the bottom of the picture.

The circle with the dot inside moves with the frame as you turn it, so after the frame has been turned and you stand back, facing it, the dot will appear to be in a different position in the circle.

Your task is to figure out exactly what the circle with the dot inside will look like in its new position, after the frame has been turned.

To understand how this works, look at the solution to Example 1 on the next page.

EXAMPLE 1:



Remember, the first step is to mentally turn the frame so that the circle with the dot inside is located at the bottom of the picture. The arrow shows that, for Example 1, the frame should be turned to the left one-quarter of a full turn.

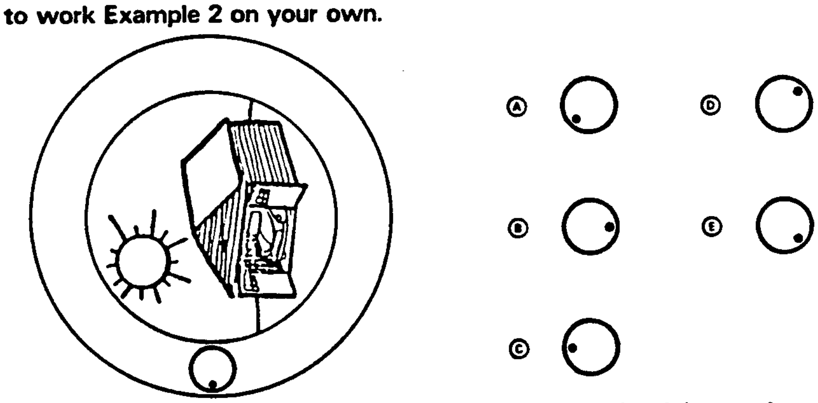
As the frame is turned, it carries with it the circle with the dot inside. So, in Example 1, after the frame has been turned, the circle with the dot inside will look like this: with the dot at the bottom of the circle. This is the answer shown in answer choice C, so the C has been darkened.

Thus, to solve each problem, you need to:

1. Mentally turn or rotate the frame until the bottom of the frame (the circle with the dot) is located at the bottom of the picture, and then
2. Figure out what the circle with the dot inside will look like in its new position.

Now try to work Example 2 on your own.

EXAMPLE 2:



For Example 2, you would turn the frame one-quarter turn to the right, so that the circle with the dot inside is located at the bottom of the picture. After you have done this, the circle with the dot inside will look like this: in its new position.

Answer B shows the circle with the dot inside in this position, so you should have darkened B.

Are there any questions about how to complete this test? Your score will be the total number of correct answers. Work as quickly and accurately as you can. You will have 10 minutes to work on this test.

DO NOT BEGIN UNTIL TOLD TO DO SO.

1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es])

Domain: Spatial abilities
Spatial Sub-domain: Spatial Relations

Orientation is the ability to keep one's bearings with respect to the environment despite changes in perspective or relative location. The items in this test also require the ability to identify pictured objects. That part of the task is intended to be very easy.

How is this construct related to other constructs/factors
For factor analytic evidence, see below 3.A.

B. Taxonomy/types/categories of items
Difficulty was varied by manipulating the amount of rotation required for correct solution.

How the test is scored/scaled? Total correct.

How are item type and difficulty jointly sampled
To be determined by examination of the items' design and statistical parameters.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers): None apparent

D. Basis for expecting incremental validity in specific jobs
ASVAB does not now measure spatial abilities. Field observations by Project A researchers suggested that maintaining one's spatial orientation is critical in Army jobs such as Tank Crew and Air Defense Artillery Crew and on tasks requiring use of maps and landmarks (i.e., in many jobs).

2. Precision of measurement

A. Test's consistency across

Different forms: Orientation has one form at present

Testing occasions: .70 in 499 incumbents

Levels of total test score (score conditional SEM):
These analyses have not yet been carried out on data from Project A. The primary analyses of interest will be against criteria of map reading, target identification, navigation, and general hands on proficiency.

B. For power or speeded test: internal consistency
In > 9,300 incumbents, Alpha = .89; split-half (Spearman-Brown) = .89

C. Variation across subgroups in score-conditional SEM

To be determined in the contract that follows up Project A.

D. If score is model-based, what's the model-based precision

N.A.

E. Subgroup differences in item functioning

Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the number correct measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	6,003	12.52	6.28
Blacks	2,553	7.79	4.46
Hispanics	334	9.72	5.73
Other	355	10.45	6.14
Females	887	9.75	5.35
Males	8,358	11.17	6.24

3. Validity

A. Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB

Spatial factor/construct: On p. 81, the six spatial tests in the CV formed the second factor in a principal factor analysis of ASVAB subtests and new Project A tests. Orientation loaded .5 on the spatial factor and .34 on the factor formed by the unspeeded tests of ASVAB. The ASVAB subtests which it correlates with most highly (> .40) are MC, MK, and AR. The squared multiple R with all ASVAB subtests is .29.

B. Uniqueness re ASVAB. How assessed?

Orientation had a uniqueness of .60 and .41 relative to internal consistency and retest reliability.

C. Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated? Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the

criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance.

Orientation correlated .242 with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + S) and .145 with AFQT. In various GLM analyses with all ASVAB and/or other Project A tests entered (p. 100), Orientation did contribute significant unique variance (p =.015 for variance not overlapping with Project A tests nor ASVAB subtests). It was the last predictor to enter significantly in the stepwise regression with all other Project A tests adding .0117 to the R^2 (p = .0453).

In 1988, a battery of four Project A tests -- One- and Two-hand Tracking, Mazes, and Orientation -- was installed in two training bases for a pilot test on students in the 11H anti-tank and 19K tank crew courses. In samples of these students (pp. 101-2), the validity of an optimally weighted combination of all ASVAB subtests against scores on TOW and anti-tank gunnery simulators was .266 and .480, respectively. Orientation correlated .162 with scores on the TOW simulator and .328 with speed/accuracy on the simulator of tank gunnery. In the 1988 samples, Orientation did not account for unique variance nor enter significantly into the stepwise regressions. By itself, it accounted for less than 1% of the residualized variance, and combining it with the other three predictors variously had little effect on R^2.

D. Subgroup differences in incremental validity or in std err of prediction of the criterion?

In 88 black and 444 white trainees in MOS 11H, TOW Gunner, score conditional SEP of accuracy in the first qualifying table (10 firings) in 1988 were:

	Black	White	p
+2 sd	6.3	1.3	.862
+1	4.2	0.8	.198
mean	2.4	0.6	.000
-1	2.1	0.9	.000
-2	3.6	1.4	.000

E. Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants: See pp. 6 - 7.

Army's combat MOS (Infantry, Armor, Field Artillery, Combat Engineer) require large numbers and are the most critical (on a scale that equates criticality with closeness to combat), but are hard to fill.

Attrition from Army training for failures of ability is rare. Instead, attrition occurs for low motivation, adjustment, fitness, discipline, and the like.

4. Equating

A. Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?) The pool of potential orientation items is unlimited. By varying the orientation of the picture, the location of the circle with the dot, and the orientation within the circle of the dot, all of which can vary continuously, we could create a very large pool. Analyses of existing data in the Project A database may reveal whether the relation of item difficulty to those variables is tight enough to make pretesting of new items unnecessary, except for small debugging trials.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed?
Instructions for Orientation make use of an enlarged sample test item, which has two pieces: the basic item and a separate, movable frame for the picture. When this test is given to large groups on paper, either more than one enlarged sample item is needed, or else the TA needs to demonstrate it from several locations in the room, so everyone can see it clearly.

C. Subgroup dependence of equating: To be determined when we get into equating.

D. Practicality of pencil-and-paper versions
The existing version is on paper.

5. Feasibility for operational S/C

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees: Ten minutes are allowed for doing the items. The present instructions for examinees are needlessly long and unclear. Administration on computer may permit the verbal instructions to be simplified by enabling presentation of more sample test items. Est. time for present instructions: 10 min. Because the test has been administered lock step to groups, no opportunity has been given for times to vary.

B. Susceptibility to

Cheating, compromise: Not certain. These dangers could be reduced by developing and using a very large item pool

Irrelevant response strategies: Individual examinees believe that they discover helpful strategies (e.g., physically rotate the page to see what the stimulus looks like after rotation. That changes the task from mental to physical rotation.). Whether any such strategies are successful remains to be seen.

Practice: In 500 incumbents, the gain was .27 s.d.

Coaching: Unknown.

C. Floor/ceiling effects? Performance of low/high AFQTs on the test? The mean of 11.02 and s.d. of 6.18 on over 9,000 cases in the concurrent validation put the mean more than two s.d. below the maximum possible score and .8 s.d. above chance.

D. Examinees' reactions: motivation, enjoyment, boredom: Nothing noteworthy

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required: No additional hardware would be required, but the graphics in the present test are likely to be unclear on a computer screen. They are already unclear on paper. That fact, and the likelihood that the ability tested does not depend much on the exact nature of the picture within the frame of each item suggest that the test could be improved by simplifying the picture. The "picture" could be an arrow, and the task then could be to rotate the frame until the circle with the dot is at the point of the arrow.

F. Need for special eqpt with special maintenance? None

Outline of the case for Reasoning to be in ECAT Validation

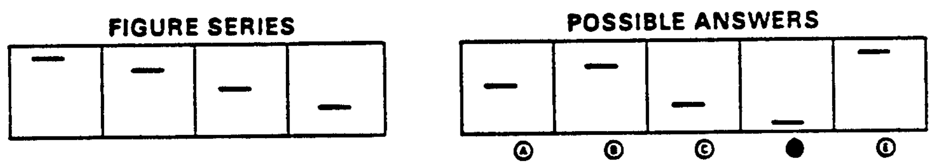
Description: Reasoning is a somewhat speeded 30-item test from the spatial battery of Project A. Given a row of four figures that form a logical series, examinee selects from multiple choices the one that is the next step in the series. Directions and a sample item are given below.

REASONING TEST

This test has rows of figures like those shown below. Each row across the page is one problem for you to solve.

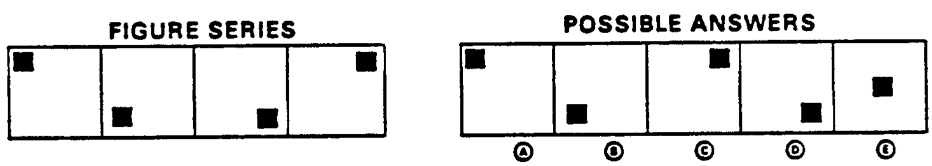
Note that each row has four figures on the left called FIGURE SERIES and five figures on the right called POSSIBLE ANSWERS. You need to discover what the pattern is in the FIGURE SERIES (what is going on as you move from one figure to the next), and then decide which one of the POSSIBLE ANSWERS would appear next in the series. Please look at Example 1.

Example 1



In Example 1, note that the lines in the FIGURE SERIES are moving lower and lower in the boxes. The first line is almost at the top, and as you go from square to square, the line sinks lower and lower. What would the next square in the series look like? Where would the line be? The line would be near the bottom of the square, as in answer Choice D. So the correct answer is D, and circle D has been filled in. Now complete Example 2 on your own.

Example 2



In Example 2, the black square in the FIGURE SERIES is moving around counterclockwise inside the larger square: it starts in the upper left corner, goes to the lower left corner, to the lower right corner, and to the upper right corner. Its next move will put it back in the upper left corner. So A is the correct answer and you should have filled in circle A.

For every problem, you must first look at the FIGURE SERIES to decide what the pattern is, and then choose from among the POSSIBLE ANSWERS the <u>one</u> that comes next in the series.

Are there any questions about how to complete this test? Your score will be the total number of correct answers. Work as quickly and as accurately as you can. You will have 12 minutes to work on this test.

1. **Theoretical development of construct and measure(s)**

A. **Definition of construct (process[es])**

Domains:

General intelligence

Inductive reasoning: This test measures the ability to generate and test hypotheses about the relationships in a series of abstract figures.

Working memory: Solving a problem requires keeping the relationships among successive figures in memory in order to find the principle that makes the figures a series, then holding the principle in memory while testing the multiple choices for their appropriateness.

How is this construct related to other constructs/ factors: Factor analyses of experts' judgments (Wing, Peterson, & Hoffman, 1984) found figural reasoning in the cognitive rather than spatial factor. Among the cognitive constructs, it clustered apart from verbal, quantitative, speed, and memory. See 3.A. below, Construct Validity.

B. **Taxonomy/types/categories of items:**

How the test is scored/scaled: Total number correct

How are item type and difficulty jointly sampled: To be determined in the next round of validation by examining item specifications and data.

C. **Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):** Probable SES related differences in exposure to posers and diagrammatic stimuli.

D. **Basis for expecting incremental validity in specific jobs:** The panel of expert judges in Project A estimated that figural reasoning would predict success on numerous tasks, like troubleshooting, inspecting, and repairing various types of systems (e.g., mechanical, electrical, fluid), detecting and identifying targets, and decoding and analyzing data. Estimated validities ranged from .18 to .41, centering around .3 for the tasks just cited.

2. **Precision of measurement**

A. **Test's consistency across**

Different forms: The test has one form that exists on two media: p/p and software. To date these alternate media of the same form have not been administered to the same examinees.

Testing occasions: .65 in > 468 incumbents

Levels of total test score (score conditional SEM): To be determined under the follow on contract to Project A.

B. **For power or speeded test: internal consistency:** Split-half (odd-even, Spearman-Brown) reliability of .87

C. **Variation across subgroups in score-conditional SEM:** To be determined under the follow on contract to Project A

D. **If score is model-based, what's the model-based precision** N. A.

E. **Subgroup differences in item functioning:** Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the number correct measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	6,002	20.41	5.15
Blacks	2,547	16.34	5.64
Hispanics	333	17.39	6.06
Other	354	18.26	5.99
Females	885	19.95	4.76
Males	8,351	19.01	5.74

3. **Validity**

A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** In the field test of the Project A battery, a pilot version of Reasoning correlated .47 and .74, respectively, with marker tests SRA Word Grouping and DAT Abstract Reasoning, which were administered concurrently to the 118 Army examinees (Peterson, 1987). The verbal nature of the stimuli in the first test was assumed to account for its lower correlation.

According to new data from the Navy (Wolfe,, unpublished, 1989), Reasoning correlates .655 with Ravens and .630 with a composite of ASVAB's power tests in a sample of 567 Navy recruits.

In factor analyses of the Concurrent Validation data, both with and without ASVAB, Reasoning loaded most heavily on the spatial factor and had a moderate loading on the factor formed by the power tests of ASVAB (pp. 81

- 84). Its highest correlations among ASVAB subtests are with AR (.51), MK (.47), and MC (.46). The squared multiple R with ASVAB was .34.

B. **Uniqueness re ASVAB. How assessed?** Its uniqueness is .53 for internal consistency and .44 for retest reliability in 9,332 incumbents.

C. **Incremental validity re ASVAB? Improvement in SD of the redicted criterion? How many other tests were considered when this one was validated?** In Marine JPM data to date, the Reasoning Test adds .031 in explained variance in knowledge test scores and .027 in supervisors' ratings in >1,100 Marines (Mayberry, unpublished, 1989).

In the 1987 sample of TOW gunners (p. 100), Reasoning was not significant in any of the GLM analyses. Its simple correlation with accuracy on Event 5 was .151 ($p = .0050$), compared with .179 for the Combat Composite (AR+CS+MC+AS) and .145 for AFQT.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?** To be calculated under the contract following up Project A.

E. **Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:** There is a great number of small fill MOS that do inspection, trouble-shooting, and repair of various systems: electrical, electronic, mechanical, and fluid. Presumably these MOS have counterparts in all of the other Services. It bears repeating, however, that measures of many predictor constructs will relate to specific tasks much better than to whole jobs. If Reasoning is a measure of fluid intelligence, there is some question as to whether the available performance measures are appropriate as criteria.

See above p. 6 - 7 for the validities and incremental validities of a composite of Project A "spatial" predictors for nine large fill MOS.

4. Equating

A. **Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?)** The potential item pool is unlimited. Pretesting would be important as item statistics cannot easily be determined rationally. Automated? Perhaps in part, but a great deal of judgment and adjustment of the products would probably be needed.

B. **Types of samples needed/acceptable for initial equating? Testing conditions needed?** Adequate range of abilities. Sufficient Ns in the small cells.

C. **Subgroup dependence of equatings:** To be calculated in pre-implementation research.

D. **Practicality of pencil-and-paper versions:** The present form is pencil-and-paper.

5. **Feasibility for operational S/C**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees:** 12 minutes is allowed for doing the items. Instruction time is about 5 min. more in the setting of a group test. NPRDC has computerized this test. Their software permits examinees to skip questions, presenting them again at the end of the test, if time allows. Dave Alderton of NPRDC reports that the computerized version of Reasoning has shown the following time characteristics in a sample of 1,305 Navy recruits recently:

Component\	mean	s.d.	maximum time
Instructions	2.59 min	1.08	12.37
Items	9.33	2.25	12.08
Total	11.92	2.73	21.63

B. **Susceptibility to:**

Cheating, compromise: Probably low

Irrelevant response strategies, coaching: Probably low

Practice: On retesting, 500 incumbents gained an average of .27 s.d., from 19.16 to 20.64 out of 36.

C. **Floor/ceiling effects? Performance of low/high AFQTs on the test?:** Floor/ceiling effects not apparent. Performance by AFQT level to be calculated in the contract that follows up Project A.

D. **Examinees' reactions: motivation, enjoyment, boredom:** Nothing noteworthy

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** None: it's already up at NPRDC on the CAT system.

F. **Need for special eqpt with special maintenance?** None

COMPUTERIZED TESTS OF PSYCHOMOTOR, PERCEPTUAL, AND MEMORIAL ABILITIES (pp. 41 - 80)

The computerized battery from Project A is a diverse set of tests that are administered on a portable computer. Stimuli and self-paced directions are shown on the computer's video screen, and the examinee responds on a special response device (p. 40). By pressing buttons on the response device the examinee registers speed and choice of response; joy sticks and sliding controls let the examinee control the movement of a crosshair on the screen. A dial is used to indicate the hand, and thus the joystick, used for two of the tests.

For measuring reaction time, the stimuli are not presented until the examinee is pressing all four "home" buttons. That constraint insures that examinees start pressing buttons from a standard hand position.

Different tests yield different scores, the variety of scores including reaction time as decision and movement speed, accuracy as proportion correct, accuracy as distance off target, and intervals between the actual and effective times to respond.

The following eight computerized tests are proposed:

Cannon Shoot	Number Memory	Short-term Memory
Two-hand Tracking	Target Identification	Target Shoot
	One-hand-tracking	

To date, validities have been reported in Project A only for a composite of the computerized predictors, not for the individual ability tests. That composite includes scores from all of the tests. The "computer composites" in the table below are six sets of non-overlapping scores from the new predictors. Those

Project A Concurrent Validation
Validities Against Hands-on Performance in Nine Major MOS
For ASVAB and Computerized Predictor Composites

	Occupations									
Predictor Composites	11B	13B	19E	31C	63B	64C	71L	91A	95B	ALL
ASVAB (4 factors, unit wtd)	.44	.22	.40	.65	.30	.54	.54	.58	.59	.47
All computer composites (unit wtd)	.42	.23	.29	.52	.25	.43	.51	.51	.44	.40
Computer + ASVAB	.49	.25	.41	.64	.34	.54	.57	.58	.60	.49

Source: McHenry, Szenas, & Wise (1987, unpublished)

Names of the occupations in the table above are:

11B	Infantryman	64C	Motor transport operator
13B	Cannon crewman	71L	Administrative specialist
19E	Armor crewman	91A	Medical specialist
31C	Radio teletype operator	95B	Military police
63B	Light-wheel vehicle mechanic		

six composites are based on the analysis on pp 81 - 85, but Complex Perceptual Speed and Accuracy is broken into two predictors, one each for speed and accuracy.

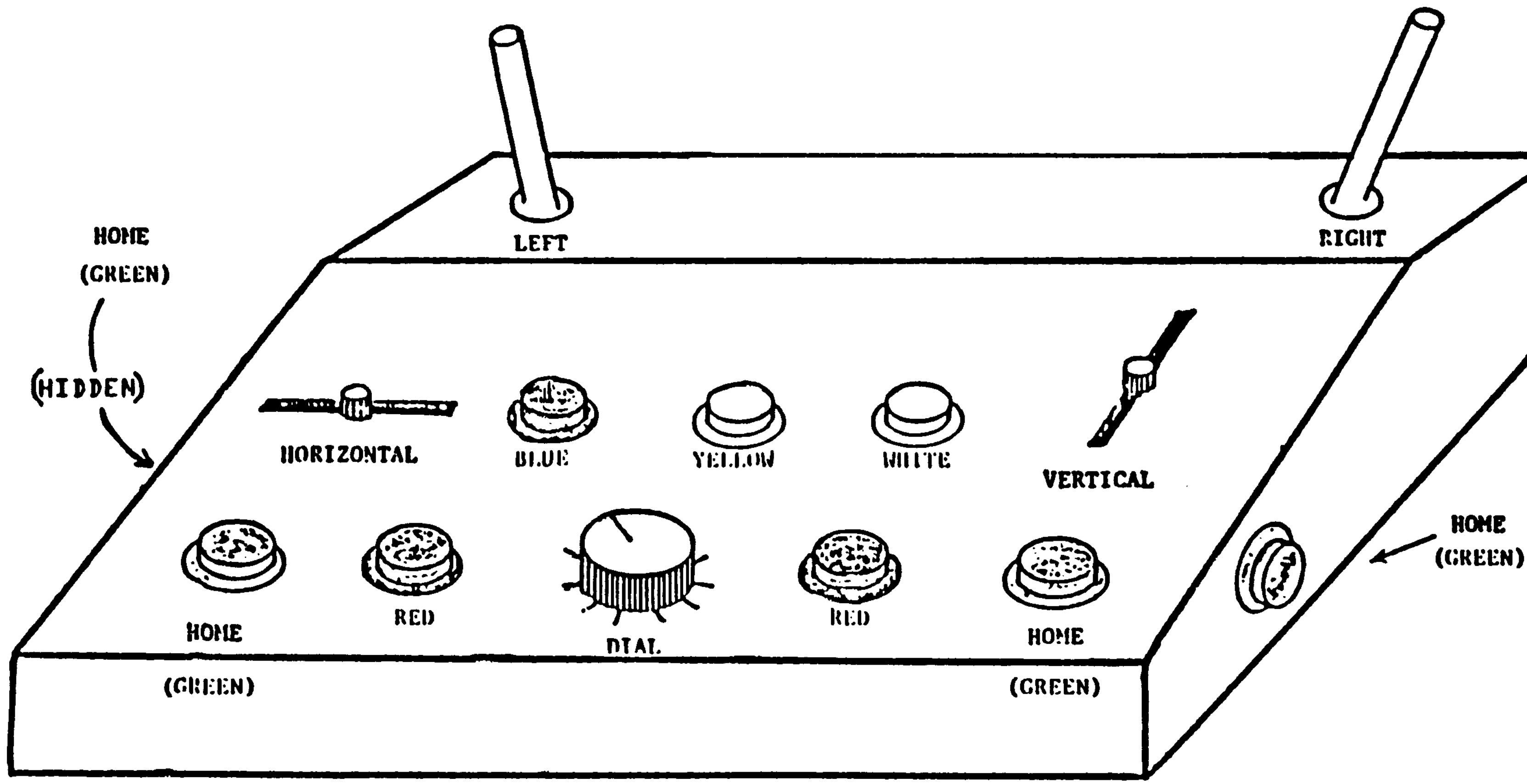
Separate and incremental validities for a unit weighted composite of those smaller composites are given in the above table. These aggregate analyses are not repeated separately for each test at 1.D and 3.C of the TASP criteria. The criterion measure is a composite of MOS-specific hands-on tasks.

In the interests of evaluating end-of-training measures, which are likely to be written tests, the validities of Project A's computerized tests for predicting written measures are shown in the next table. The predictor is a composite of the new computerized tests; the criterion is a composite of end-of-training and on-the-job knowledge tests in the CV. Unlike end-of-training tests that are a part of training courses, the Project A knowledge tests were designed to have ample variances. Thus, the correlations in this next table are probably an optimistic estimate of the validities to be expected in the ECAT validation.

Project A Concurrent Validation
Validities Against Knowledge Tests in Nine Major MOS
For ASVAB and Computerized Predictor Composites

Predictor Composites	11B	13B	19E	31C	63B	64C	71L	91A	95B	ALL
ASVAB (4 factors, unit wtd)	.77	.61	.76	.77	.80	.74	.73	.81	.78	.75
All computer composites (unit wtd)	.65	.59	.69	.68	.59	.63	.65	.66	.67	.65
Computer + ASVAB	.78	.64	.78	.77	.80	.77	.75	.81	.79	.77

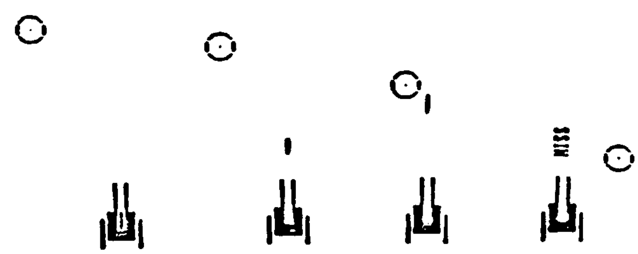
Source: McHenry, Szenas, & Wise (1987, unpublished)



Response device for Project A computerized predictors

The case for Cannon Shoot to be in the ECAT Validation

Description: Cannon Shoot is a 36-item test from the computerized battery of Project A that involves estimating the time to fire a projectile from a static cannon so as to hit a target moving across the line of fire. The target and shell move in straight lines. Speed of the shell is constant across trials; target speed and direction vary over trials. Practice items and self-paced instructions are given on the screen. A sequence of four sample screens is given below.



1. **Theoretical development of construct and measure(s)**

A. Definition of construct (process[es])

This task involves judgment of rates of visible movement and memory for the rate of the cannon shell. Peterson (1987) referred to the construct as "estimating relative velocities."

How is this construct related to other constructs/factors
See below 3.A., Construct Validity

B. Taxonomy/types/categories of items
Items vary in the angle of target movement relative to the cannon (12 possible values), distance from the cannon to the point of impact (4 possible values), and distance the target travels from time of firing to time of direct hit (4 possible values).

How the test is scored/scaled:
The following scores have been recorded at various stages in Project A: mean minimum distance between shell and target (raw and log transformation), mean times between actual and optimal firing (absolute, for early shots alone, and for late shots alone), and proportion of hits (hit = shell passes within the boundaries of the target). Time measures are in hundredths of a second; logs are natural logs.

How are item type and difficulty jointly sampled?
Items' design parameters all affect their difficulty. For the field test, a Latin square design was used to generate 48 items in the matrix of possible combinations of parameter values. The field test battery was reduced by 12 items. In the remaining 36 some are easy enough to have little variance, but the distribution of difficulty across items remains to be checked.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers): In the field test (p.), a correlation of .18 ($p < .05$) between test scores and self-reports of experience with video games was found.

D. Basis for expecting incremental validity in specific jobs. It appears to researchers in Project A that judgment of movement is involved in visually guided gunnery, driving (trucks, tanks, and construction equipment), air traffic control, and calling for fire on moving targets. ASVAB, of course, does not measure this construct. This construct was not included in the expert judgment task which guided the choice of predictors in Project A.

2. **Precision of measurement**

   A. **Test's consistency across**

   Different forms: The test has one form at present.

   Testing occasions: For the measure of mean absolute time between actual and optimal firing, reliability over an interval of two weeks was .52 in 480 incumbents. An alternative scoring method, one that discards items which have low variance, is to be evaluated soon in an effort to improve retest reliability.

   Levels of total test score (score conditional SEM): To be determined in the follow on contract to Project A.

   B. **For power or speeded test: internal consistency:** An odd-even reliability of .65 was observed on the 9,000+ cases in the CV.

   C. **Variation across subgroups in score-conditional SEM:** To be determined in the follow on contract to Project A.

   D. **If score is model-based, what's the model-based precision** N.A.

   E. **Subgroup differences in item functioning:** Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the latency measure (mean discrepancy between optimal and actual time to fire) by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,992	4.93	13.52
Blacks	2,542	7.23	15.92
Hispanics	334	4.46	13.81
Other	354	4.96	12.85
Females	885	5.52	18.33
Males	8,337	5.55	13.74

3. **Validity**

   A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** Although it involves minimal motion (viz., pressing a button that the hand is poised over), Cannon Shoot's mean time discrepancy loads most heavily with psychomotor scores (pp. 81 - 84). The highest correlation with an ASVAB subtest is .30 for MC, while the median correlation with ASVAB subtests is .19 (p. 87). Squared multiple R with all ASVAB subtests is .09 for that same score.

B. **Uniqueness re ASVAB. How assessed?** Uniqueness relative to all ASVAB subtests is estimated at .56 and .43 for the mean absolute time discrepancy score, using internal consistency and retest reliabilities, respectively.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, taken the Project A tests during their first three days of Service, and then been tested on their anti-tank gunnery skills about twelve weeks later. The criterion measure was an accuracy score on the fifth event (i.e., set of 10 trials), which was the first event on which a gunner could attain qualifi cation. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance (p. 100).

Cannon Shoot correlated .159 (.0029) with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. The multiple R^2 for all ASVAB subtests was .041, for all Project A was .174, and for both sets together .181. Project A tests accounted for .133 of the residual variance. In the various GLM analyses, Cannon Shoot did not account for significant unique variance, nor enter into the stepwise regression.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?** To be determined under the contract that follows up Project A.

E. **Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:** Validity is more likely to generalize across jobs at the level of [conceptually appropriate] tasks than at the level of the whole job. The gunnery tasks requiring movement judgment are in large fill, hard to fill MOS. Likewise for tank driving. The MOS for truck drivers has fairly large numbers. Air traffic controllers are not numerous, but utility of good performance is high.

4. Equating

A. Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated? As there are several continuous dimensions of item difficulty, the potential item pool is large. Equivalent forms could be produced by merely scrambling or rotating the existing items. An algorithm to generate items could be written.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed?
Numbers of minority and low ASVAB examinees will set lower limits on sample sizes.

C. Subgroup dependence of equatings: To be determined.

D. Practicality of pencil-and-paper versions
Nil. Computerization makes this kind of test possible.

5. Feasibility for operational S/C

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees: The correlation of time to read the instructions for the whole computerized battery and AFQT is about .30. In the CV, mean total test time plus and minus two standard deviations equalled 10m 09sec to 3m 33sec.

B. Susceptibility to:

Cheating, compromise, irrelevant response strategies: Probably low. The latency variable is reckoned from the optimal time to shoot, so there is not a speed/accuracy tradeoff.

Practice: On retesting, 487 incumbents improved their accuracy by .3 s.d. on the time discrepancy score (p. 98)

Coaching: Unknown; doubtful

C. Floor/ceiling effects? Performance of low/high AFQTs on the test? To be determined under the contract that follows up Project A

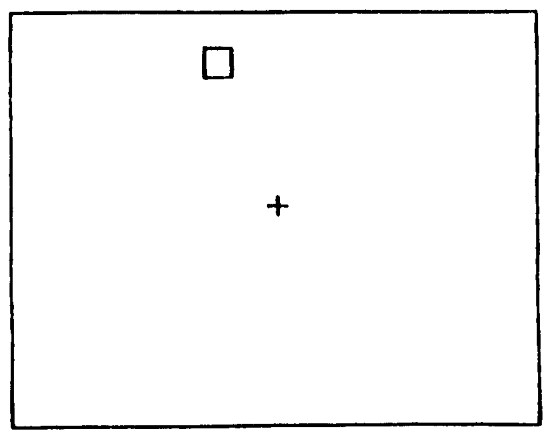
D. Examinees' reactions: motivation, enjoyment, boredom: Fun

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required: Any single key could serve to fire the cannon, so no special response device is needed.

F. Need for special eqpt with special maintenance?
None.

Target Shoot

Target shoot is a 30-trial test of psychomotor abilities from the computerized tests of Project A. The examinee uses a joy stick to move a crosshair inside a target that moves unpredictably on the video screen, and then shoot the target by pressing a button with the other hand. As soon as the examinee fires, the trial ends. Practice items and self-paced instructions are given on the screen. A sample screen is shown below.



1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es])

Domain: Psychomotor

This construct is the ability to make movements so as to adjust a movable part accurately. For this test, the adjustment is made in response to erratic, rather than predictable, movements of the stimulus. Fleishman's psychomotor constructs of control precision and rate control are involved.

How is this construct related to other constructs/factors
Control precision is involved in the 1-hand and 2-hand Tracking tests as well, but Target Shoot involves the added ability of adjusting to unpredictable movement in the stimulus, which Fleishman defines as Rate Control. In addition, the examinee must make a gross movement with the other hand at just the right moment, so as to hit the target. Precision in "firing" may involve aiming ability, quickness of reaction, and/or multi-limb coordination, all of which are constructs of Fleishman's.

B. Taxonomy/types/categories of items:
Items vary in difficulty with average target speed, maximum crosshair speed, and frequency of changes in course and speed by the target.

How the test is scored/scaled?
The main measure of interest is Mean Log (Distance + 1), which is distance off target at the moment of firing, percent hits, and time to fire.

How are item type and difficulty jointly sampled?
This question will be answered as soon as we can examine the item specifications and data.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):
Self reported experience with video games correlated .27 ($p << .05$) with Target Shoot for 256 incumbents in the field test of Project A. Such experience may related to SES. Content would offend only those who disapprove of shooting.

D. Basis for expecting incremental validity in specific jobs
- Face validity: this test resembles in two dimensions a real gunnery task
- The judgment of experts in Project A (Wing, Peterson, and Hoffman, 1984) was that measures of rate control and precision control would be good predictors of all kinds of gunnery: heavy direct fire weapons,

individual weapons, and indirect fire.

- ASVAB does not have psychomotor tests; and psychomotor and general cognitive abilities have a relatively low correlation

2. **Precision of measurement:**

A. **Test's consistency across:**

Different forms: Target shoot has one form at present.

Testing occasions: Retest reliabilities of .37 and .58 were found for the error and time to fire scores, in order, in 480 incumbents.

Levels of total test score (score conditional SEM): To be computed in the follow-on to Project A.

B. **For power or speeded test, internal consistency:** The odd-even reliabilities of the error and time to fire measures were .74 and .85 in 9,000 incumbents.

C. **Variation across subgroups in score-conditional SEM:** To be calculated in the follow on to Project A.

D. **If score is model-based, what's the model-based precision** N. A.

E. **Subgroup differences in item functioning:** Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the error measure, mean log (distance off target at moment of firing, +1), by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,870	2.16	0.23
Blacks	2,346	2.22	0.26
Hispanics	324	2.15	0.20
Other	340	2.18	0.26
Females	774	2.31	0.29
Males	8,106	2.16	0.23

3. **Validity**

A. **Construct validity: correlations with other tests of similar things. Intercorrelations with ASVAB:** In a factor analysis with pre-enlistment ASVAB subtests and the Project A spatial and computerized predictors (p. 81), the error and time measures loaded most heavily on the psychomotor factor. Other factor analyses with fewer predictors provide similar clustering and loadings (pp. 82 - 4). The loadings on the psychomotor factor are moderate, and no strong secondary loadings are present.

Target shoot has low correlations with all ASVAB subtests (median < .14, max. .23) (p. 90), and a squared multiple correlation with all the subtests < .10 (p. 95).

B. **Uniqueness re ASVAB. How assessed?** Estimated at .70 and .37 for the error score with respect to internal consistency and retest reliability.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, taken the Project A tests during their first three days of Service, and then been tested on their anti-tank gunnery skills about twelve weeks later. The criterion measure was an accuracy score on the fifth event (i.e., set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - were was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance (p. 100).

Target shoot correlated .112 with the criterion, compared with .179 for the Combat Composite and .145 with AFQT. The multiple R^2 for all ASVAB subtests was .041, for all Project A was .166, and for both sets together .177. In predicting the residual criterion variance, Target Shoot did not contribute unique variance, and it did not enter significantly into the stepwise regression.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?** To be computed in the follow-on to Project A.

E. **Job families and patterns of validities for them Similarity of job families in other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:** The job family or whole single job is probably too gross a unit of job analysis to display a single test's potential for helping with classification. Target Shoot is expected to show gains in validity against gunnery tasks, which are critical in the large fill, hard to fill combat MOS.

4. Equating

A. **Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?)** The potential item pool is very large. Other forms could be created by scrambling the order and/or spatial orientation of the present items. Items probably could be created by an algorithm, but writing the algorithm would cost something. Would pretesting ever not be advisable?

B. **Types of samples needed/acceptable for initial equating?** Samples with adequate numbers of the small subgroups

Testing conditions needed? Suitable for testing in general and computerized testing in particular. Nothing unusual

C. **Subgroup dependence of equatings:** To be addressed when equating reached.

D. **Practicality of pencil-and-paper versions:** Zero; this kind of test is possible because of computers.

5. **Feasibility for operational S/C:**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees:** Mean total test time was 4m 50s in 9,208 incumbents (s.d., 30s). Times have not been broken out by AFQT level, but the correlation of time to read the directions on the computerized tests in Project A with AFQT is about .30.

B. **Susceptibility to:**

Cheating, compromise: Given the large potential set of items, the danger here may be small.

Irrelevant response strategies: To counter a strategy of sacrificing accuracy for speed, either use an accuracy score, or develop a composite predictor that adjusts speed of firing for miss distance. In principle, an examinee could try to hit near the target, instead of right inside. Whether that strategy would affect the rank order of scores is not yet known.

Coaching: Unknown; not obviously large.

Practice: Gains on a first retesting were negligible in > 472 incumbents (p. 96).

C. **Floor/ceiling effects? Performance of low/high AFQTs on the test?** Variability is great enough to produce significant correlations. Easiness is not a problem:

the mean log (distance + 1) is about 9 s.d. above the theoretical perfect level, and mean time to fire is about 5 s.d. above instantaneous firing.

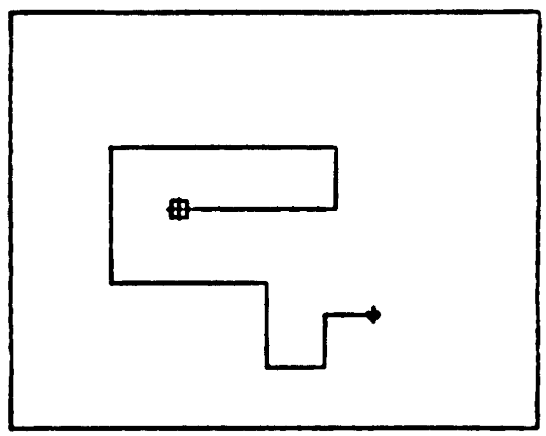
D. **Examinees' reactions: motivation, enjoyment, boredom:** This test is fun.

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** Something like the Project A response pedestal would be needed: at least two joy sticks and two buttons (for selecting the right or left joy stick).

F. **Need for special eqpt with special maintenance?** None other than that above

The case for One-hand Tracking to be in the ECAT Validation

Description: One-hand tracking is an 18-item test of eye-hand coordination from the computerized battery of Project A. The examinee uses a joy stick to try to keep a crosshair inside a square as it moves along a path. The path, which is explicitly marked on the video screen, consists of straight horizontal and vertical segments. Practice items and self-paced instructions are given on the screen. A sample screen is shown below.



Theoretical development of construct and measure(s)

A. **Definition of construct (process[es]):**
This test measures Fleishman's precision of control in the psychomotor domain. Control precision is "the ability to make fine, highly controlled movements to adjust a machine control mechanism in response to a stimulus whose speed and direction of movement are perfectly predictable. (p. 5-39)" (Peterson, 1987)

How is this construct related to other constructs or factors? Tasks involving the ability of multi-limb coordination may, like the Two-hand Tracking test in Project A, also involve control precision. Fleishman's factor of rate control is defined in part by the unpredictable path of the target, so it is not involved here.

B. **Taxonomy/types/categories of items:**
Items vary in difficulty with maximum speed of the cross-hairs, difference between that speed and target speed, and number of turns. Target speed and total length of path are controlled together so that the total time for each item, and thus the number of measurements of distance off target, would be equal for all items.

How the test is scored/scaled? The one score which has been used for analyses to date is Mean Log (Distance off Target + 1)

How are item type and difficulty jointly sampled? The item parameters are crossed to produce items of varying difficulty, then the sequence of items is determined by randomly sampling the matrix. This procedure was chosen on the basis of evidence in pilot tests that there was little practice effect within a testing session.

C. **Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):** Items are content-free, so offensiveness if not an issue. There may be SES-related differences in exposure to video games. Self-reported exposure correlated .22 with One-hand Tracking in 256 incumbents in the Project A field test ($p << .05$).

D. **Basis for expecting incremental validity in specific jobs:** In the expert judgment exercise early in Project A, 35 judges estimated the validity of control precision against gunnery tasks - direct fire of heavy weapons, indirect fire, and use of individual weapons - to be .25 to .29. ASVAB is conceptually and statistically

independent of psychomotor measures, so the potential for incremental validity is good.

2. **Precision of measurement**

   A. **Test's consistency across**

   Different forms: One form of the test exists now. The intercorrelation of One- and Two-hand Tracking give a rough estimate of alternate form reliability of One-hand Tracking, since both use the same screens. That figure ranges from .77 in the Concurrent Validation of Project A in 1985 to .81 in both TOW gunnery trainees and tank crew trainees in 1988.

   Testing occasions: Retest reliability was .74 in 468 first tour soldiers.

   Levels of total test score (score conditional SEM): To be computed under the follow on contract to Project A.

   B. **For power or speeded test: internal consistency:** Odd-even reliability is .98 in the CV sample of 9,000+

   C. **Variation across subgroups in score-conditional SEM:** To be computed under the follow on contract to Project A.

   D. **If score is model-based, the model-based precision:** N. A.

   E. **Subgroup differences in item functioning:**

   Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the error (distance) measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,999	2.87	0.44
Blacks	2,550	3.23	0.50
Hispanics	336	3.02	0.48
Other	354	3.04	0.52
Females	883	3.39	0.46
Males	8,356	2.94	0.47

3. **Validity**

   A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** As noted above, the correlation of One- and Two-hand Tracking is in the .77 to .81 range. In factor analyses of Project A tests with and without pre-enlistment ASVAB (pp. 81 - 85), One-hand Tracking anchors the psychomotor factor and has no

strong loadings elsewhere. Its median intercorrelation with ASVAB subtests is .235, the highest correlations being with AS (.30), GS (.28) and EI (.27).

B. **Uniqueness re ASVAB. How assessed?** In terms of internal consistency and retest reliability, uniqueness is estimated to be .82 and .58, respectively.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance.

One-hand Tracking correlated .206 with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. In various GLM analyses with all other Project A tests entered (p. 100), One-hand Tracking did not account for significant unique variance.

In 1988, a battery of four Project A tests -- One- and Two-hand Tracking, Mazes, and Orientation -- was installed in two training bases for a pilot test on students in the 11H anti-tank and 19K tank crew courses. In samples of these students (pp. 101-2), the validity of an optimally weighted combination of all ASVAB subtests against scores on TOW and anti-tank gunnery simulators was .266 and .480, respectively. One-hand Tracking had a validity of .290 and .484. In the 1988 samples, One-hand Tracking did not account for significant unique variance. But by itself, it accounts for .054 of the residualized variance, and combining it with Mazes (R^2 = .007) increases the residual variance accounted for to .076.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?**
Against the criterion of accuracy in firing the M70 TOW simulator in 1988, in SD units of Mean Log (Distance + 1) on the composite of 1- & 2-hand Tracking:

		-2	-1	0	+1	+2
SEP W	(n = 457)	1.4	0.9	0.6	0.8	1.3
SEP B	(n = 89)	2.9	2.0	2.1	3.2	4.6

E. **Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:**
TASP has agreed that the focus of the new tests will be classification. It seems likely that the gains in validity over ASVAB for any new test will be more obvious at some level of job analysis finer than the whole job. For One-hand Tracking, conceptually appropriate tasks include

- Visually guided gunnery, which is critical in Infantry, Tank Crew, some Air Defense jobs,

- Driving, which is a separate job, and is also involved in Combat Engineering, Mechanized Infantry, Tank Crew

Infantry and Tank Crew are large fill, combat critical MOS which, because of the lack of similar civilian jobs, are hard to fill. Army spends large sums recruiting to fill those specialties.

4. Equating

A. **Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item development be automated?** Equivalent forms would be easy to develop, either from the limitless population of potential new items or from scrambling existing items. Item development could be done by algorithm. There may not be enough items in existence now to allow predicting empirical parameters of new items.

B. **Types of samples needed/acceptable for initial equating? Testing conditions needed?** The need for enough minorities and low AFQT Examinees would drive sample size.

C. **Subgroup dependence of equatings:** To be determined when equating analyses are done.

D. **Practicality of pencil-and-paper versions:** The primary outcome measure is sampled every hundredth of a second, so it would be hard to replicate on paper without fancy apparatus.

5. **Feasibility for operational S/C**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees.** Mean total time for instructions plus items was 6m 01s (s.d. 34s) for 9,200+ incumbents (p. 104). Times by AFQT TBD under the contract that follows up Project A.

B. **Susceptibility to**

Cheating, compromise: Probably low

Irrelevant response strategies: An examinee could try to keep the crosshair near the target instead of following the instruction to keep it on the target. This other strategy is not "irrelevant" since it would involve tracking. Whether it would produce results that are different from the instructed strategy is an empirical question.

Coaching: There may be coachable but relevant response strategies (things like "Keep your eye on the ball").

Practice: A gain of .24 s.d. at retest was found in 480 incumbents.

C. **Floor/ceiling effects? Performance of low/high AFQTs:** Mean distance off target has a large maximum score. Scores by AFQT TBD in the follow-on to Project A.

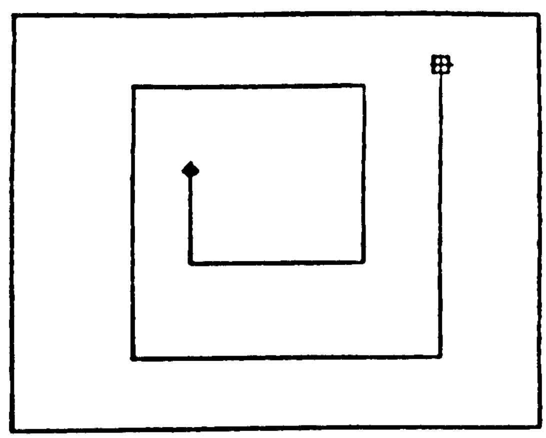
D. **Examinees' reactions: motivation, enjoyment, boredom:** This test is exciting and difficult.

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** Double digital to analog game board, clock, and response pedestal with two joy sticks.

F. **Need for special eqpt with special maintenance?** The buttons and controls on the response pedestal sometimes need repair or replacement.

The case for Two-hand Tracking to be in the ECAT Validation

Description: Two-hand tracking is a an 18-item test of multi-limb coordination from the computerized tests of Project A. The examinee uses two sliding controls, one moving horizontally and one vertically, to try to keep a crosshair inside a square as it travels along a path. The path, which consists of straight vertical and horizontal segments, is explicitly marked on the video screen. Practice items and self-paced instructions are given on the screen. A sample screen is shown below.



1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es])

This test measures the ability to coordinate the simultaneous movement of two limbs. Because the movement is to adjust a crosshair to track a target moving continuously on a known path, the ability of control precision is involved as well.

How is this construct related to other constructs/ factors: In a factor analysis of experts' judgments of validities of 53 predictor constructs against 72 Army performance constructs (Wing, Peterson, & Hoffman, 1984), psychomotor abilities formed a factor containing clusters of dexterity, coordination, and precision/ steadiness. The judged constructs forming the latter two clusters include multi-limb coordination and control precision. For the relation of scores on this test to scores on the others others in the Project A battery and to ASVAB, see below Construct Validaty, 3.A.

B. Taxonomy/types/categories of items:

Items vary in difficulty with maximum speed of the cross-hairs, difference between that speed and target speed, and number of turns. Target speed and total length of path are controlled together so that the total time for each item, and thus the number of measurements of distance off target would be equal for all items.

How the test is scored/scaled? Mean Log (Distance off Target + 1)

How are item type and difficulty jointly sampled? The item parameters are crossed to produce items of varying difficulty, then the sequence of items is determined by randomly sampling the matrix. This procedure was chosen on the basis of evidence in pilot tests that there was little practice effect within a testing session.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers): In the field test of Project A, self-reports of 256 incumbents' experience with video games correlated .16 ($p < .05$) with their scores on Two-hand Tracking. SES may relate to differences in exposure to video games and diagrammatic graphics. Items are content-free, so offensiveness should not be an issue.

D. Basis for expecting incremental validity in specific jobs

- ASVAB does not measure psychomotor skills
- At least some tasks for many jobs appear to require psychomotor skills.

- In Project A (Wing, Peterson, & Hoffman, 1984), 35 subject matter experts estimated the true validity of multi-limb coordination against nine psychomotor tasks from Army jobs to be a median (across those tasks) of .26 (range .18 - .37).

2. Precision of measurement

A. Test's consistency across

Different forms: One form of the test exists now. The intercorrelation of One- and Two-hand Tracking give a rough estimate of alternate form reliability of Two-hand Tracking, since both use the same screens. That figure ranges from .77 in the Concurrent Validation of Project A in 1985 to .81 in both TOW gunnery trainees and tank crew trainees in 1988.

Testing occasions: Retest reliability was .85 in 487 incumbents from a mixture of MOS

Levels of total test score (score conditional SEM): To be computed under the follow up to Project A.

B. For power or speeded test: internal consistency: Odd-even reliability of .98 for 9,200+ incumbents

C. Variation across subgroups in score conditional SEM:

D. If score is model-based, what's the model-based precision? N.A.

E. Subgroup differences in item functioning: Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the distance off target measure by group in the Concurrent Validation were as follows:

Group	n	x	s.d.
Whites	5,991	3.56	0.49
Blacks	2,548	3.99	0.43
Hispanics	336	3.76	0.46
Other	352	3.77	0.51
Females	879	4.11	0.38
Males	8,348	3.65	0.51

3. Validity

A. Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB: In a factor analysis with pre-enlistment ASVAB subtests, the Project A spatial tests, and the other Project A

computerized predictors (p. 81), 2-hand Tracking (error score) had the second highest (after 1-hand Tracking error) loading on the psychomotor factor. The median of its intercorrelations with ASVAB subtests was .255, the highest correlations being with MC (.41), AS (.33), GS (.31), and EI (.30). With the 1-hand Tracking error score the correlation was .77 in the Concurrent Validation (p. 90) and .81 in the 1988 analyses of data from simulators of TOW gunnery and tank gunnery.

In PROC GLM analyses of the gunnery battery (viz., 1- and 2-hand Tracking, Mazes, and Orientation) against both TOW and tank gunnery (pp. 101-2), 2-hand tracking was the single strongest predictor of gunnery, and 1-hand tracking did not significantly enter the equations after 2-hand. These results do not support the speculation that 1-hand tracking is interchangable with 2-hand as a predictor.

B. <u>Uniqueness re ASVAB. How assessed?</u> .79, in terms of odd-even reliability minus the .19 squared multiple R with all ASVAB subtests. In terms of retest reliability (.85), .66.

C. <u>Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?</u> Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance. The results are shown on p. 100.

Two-hand Tracking correlated .294 with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. The multiple R^2 for all ASVAB subtests was .041, for all Project A was .165, and for both sets together .177. Project A tests accounted for .126 of the residual variance. In all combinations of tests, Two-hand Tracking accounted for significant

amounts of unique variance, and it entered first into the stepwise regression (R^2 = .0725).

In analyses of tank and anti-tank gunnery that include all ASVAB subtests and four new Army tests (One- and Two-hand Tracking, Mazes, and Orientation), Silva (1989, unpublished) (pp. 101-2) has found that Two-hand tracking is the single best predictor on several grounds: highest zero order correlation, significant and largest unique variance, and first of Project A tests to enter stepwise regressions predicting variance after ASVAB-related variance is removed.

D. <u>Subgroup differences in incremental validity or in std err of prediction of the criterion? See above 2.C</u>. Against the criterion of accuracy in firing the M70 TOW simulator in 1988, in SD units of Mean Log (Distance + 1) on the composite of 1- & 2-hand Tracking:

	-2	-1	0	+1	+2
SEP W (n = 457)	1.4	0.9	0.6	0.8	1.3
SEP B (n = 89)	2.9	2.0	2.1	3.2	4.6

E. <u>Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:</u> TASP has agreed that the focus of the new tests will be classification. It seems likely that the gains in validity over ASVAB for any new test will be more obvious at some level of job analysis finer than the whole job. For Two-hand Tracking, conceptually appropriate tasks include

- Visually guided gunnery, which is critical in Infantry, Tank Crew, some Air Defense jobs,

- Driving, which is a separate job, and is also involved in Combat Engineering, Mechanized Infantry, Tank Crew

Infantry and Tank Crew are large fill, combat critical MOS which, because of the lack of similar civilian jobs, are hard to fill. Army spends large sums recruiting to fill those specialties.

4. <u>Equating</u>

A. <u>Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?</u> Equivalent forms would be easy to develop, either from the limitless population of potential new items or from scrambling existing items. Item develop-

ment could be done by algorithm. It is an open question whether there are enough items in existence now to allow anticipating how new items would behave.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed? The need for enough minorities and low AFQT Examinees would drive sample size.

C. Subgroup dependence of equatings: To be determined when equating analyses are done.

D. Practicality of pencil-and-paper versions: The primary outcome measure is sampled every hundredth of a second, so it would be hard to replicate on paper.

5. Feasibility for operational S/C

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees. Mean total time for instructions plus items was 6m 01s (s.d. 34s) for 9,200+ incumbents (p. 104). Times by AFQT TBD under the contract that follows up Project A.

B. Susceptibility to

Cheating, compromise: Probably low

Irrelevant response strategies: An examinee could try to keep the crosshair near the target instead of following the instruction to keep it on the target. This other strategy is not "irrelevant" since it would involve tracking. Whether it would produce results that are different from the instructed strategy is an empirical question.

Coaching: There may be coachable but relevant response strategies (things like "Keep your eye on the ball").

Practice: A gain of .24 s.d. was found in 480 incumbents

C. Floor/ceiling effects? Performance of low/high AFQTs: Mean distance off target has a large maximum score. Scores by AFQT TBD in the follow-on to Project A.

D. Examinees' reactions: motivation, enjoyment, boredom: This test is exciting and difficult.

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required: Double digital to analog game board, clock, and response pedestal with two sliding controls.

F. <u>Need for special eqpt with special maintenance?</u> The buttons and controls on the response pedestal sometimes need repair or replacement.

The case for Number Memory to be in the ECAT Validation

Description: Number Memory is a 28 item mental arithmetic test from the computerized battery of Project A. The examinees are shown a number, then at their own pace, bring on a new screen with directions for an arithmetic operation to be performed using another number (e.g., Add 5). After a series of such operations and numbers, examinee presses a button to indicate if a probe number is the answer to the series of operations. Below are three successive screens printed together.

START WITH 14

DIVIDE BY 7

MULTIPLY BY 8

NO (BLUE) YES (WHITE)

1. **Theoretical development of construct and measure(s)**

   A. **Definition of construct (process[es]):**
   Cognitive abilities, quantitative: ability to perform the four basic arithmetic operations quickly and accurately. The format of this task requires the examinee to use working memory as well.

   How is this construct related to other constructs/ factors?

   B. **Taxonomy/types/categories of items:** Items vary in the number of steps presented before a solution is required (2, 3, or 4). The four arithmetic operations were presented equally often.

   How the test is scored/scaled: A variety of speed and accuracy measures are available, both for whole problems and for each of the four arithmetic operations. The measure which is used for analyses below is the pooled mean operation time. Pooling is within arithmetic operations across problems, then averaging is across operations.

   How are item type and difficulty jointly sampled: In a balanced fashion.

   C. **Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):** There is no content, except arithmetic. Scores on this test did not significantly correlate with self reports of video game experience in 256 incumbents in the Project A Field Test.

   D. **Basis for expecting incremental validity in specific jobs?** The hypothesis that administering a test like NO by computer could improve its statistical characteristics and validity. In the expert judgment task of Project A, Number Memory had an average mean expected validity of .40. Technical and Clerical MOS were expected to benefit the most from this ability.

2. **Precision of measurement**

   A. **Test's consistency across**

   Different forms: One form exists now

   Testing occasions: .73 for mean pooled decision time in 473 cases

   Levels of total test score (score conditional SEM): To be determined in the contract following up Project A

B. **For power or speeded test: internal consistency:** Odd-even reliability of .93 for pooled mean decision time samples of at least 7,700 incumbents.

C. **Variation across subgroups in score-conditional SEM:** To be determined in the contract following up Project A

D. **If score is model-based, what's the model-based precision?** N. A.

E. **Subgroup differences in item functioning:** Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the mean pooled decision time by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,930	222.96	75.21
Blacks	2,485	252.78	83.34
Hispanics	324	254.95	80.48
Other	350	245.36	94.47
Females	8,210	232.64	80.45
Males	879	237.60	72.62

3. Validity

A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** In factor analyses with ASVAB, the latency measures of Number Memory loaded on a quantitative factor that included AR and MK (p. 81). In other factor analyses (pp. 82-5), the company it kept varied across a mixture of quantitative and speed. Its lowest correlations with pre-enlistment ASVAB subtests were with MC and EI (p. 87).

B. **Uniqueness re ASVAB. How assessed?** In terms of internal consistency and retest reliability, uniqueness relative to pre-enlistment ASVAB was .66 and .48 for the pooled mean decision time measure.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project

A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance.

Number Memory correlated .069 ($p = .195$) with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. In the analyses with all ASVAB and/or other Project A tests entered that are summarized on p. 100, it did not contribute significant unique variance. However, in an earlier stepwise regression, before the specification of the predictor scores had been cleaned up, Number Memory entered significantly (second variable to enter, adding R^2 of .0116 ($p = .0491$)). The difference between the earlier analysis and the one on p. 100 was this: the score for Perceptual Speed and Accuracy was changed from percent correct to latency, and Number Memory was changed from mean trimmed latency of final response to mean latency of the component responses to the problems. The latter score for Number Memory seems preferable on the grounds that it captures more information, and the small sample size invites an explanation of sampling error for any unreplicated finding. Although its mean final response time entered significantly in the earlier stepwise analysis, the simple correlation of the non-preferred score with the criterion was .0153 (n.s.). In the two stepwise regressions where Number Memory behaved differently, the set of other predictors that entered significantly remained the same.

D. **Subgroup differences in incremental validity or in std err of prediction of the criterion?** To be determined under the contract that follows up Project A.

E. **Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants:** Number memory should have utility in general, across all MOS, as the quantitative, speed, and accuracy attributes that it is related to are generally useful.

4. **Equating**

A. **Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?)** The potential item pool is very large. Item development

probably could be automated.

B. **Types of samples needed/acceptable for initial equating? Testing conditions needed?** Numbers in the subgroups would be the limiting condition. No unusual testing conditions.

C. **Subgroup dependence of equatings:** To be determined when we reach the stage of equating.

D. **Practicality of pencil-and-paper versions:** Computerization makes the variety of response time measures possible.

5. **Feasibility for operational S/C**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees:** The mean and s.d. of total test time (instructions + items) were 9m 32s and 2m 26s in 9,000+ incumbents.

B. **Susceptibility to**

Cheating, compromise, irrelevant response strategies, coaching: Judged to be low

Practice: In 473 incumbents, the gain in mean pooled decision time was .44 s.d. on retesting. Mean decision time showed a gain of .26.

C. **Floor/ceiling effects? Performance of low/high AFQTs on the test?** The maximum latency is unlimited. How close to a minimum the mean latency of 2.0 secs on retesting is remains to be seen. Performance by AFQT level remains to be determined in the contract following up Project A.

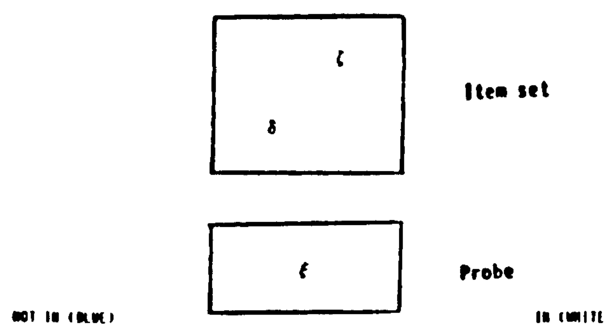
D. **Examinees' reactions: motivation, enjoyment, boredom:** Nothing noteworthy.

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** None

F. **Need for special eqpt with special maintenance?** None

The case for the Short-term Memory Test to be in the ECAT Validation

Description: Short-term Memory is a 36-item test from the computerized battery of Project A that involves judging whether a single symbol on the video screen was in an array of symbols (English letters, Greek letters, printers' symbols, etc.) that just disappeared. In order for the stimuli to be presented, the examinee's hands must be in a standard position on the reponse device. Practice items and self-paced instructions are given on the screen. A sample screen is shown below.



1. **Theoretical development of construct and measure(s)**

A. **Definition of construct (process[es]):**
Speed and accuracy with which one searches, recalls, and judges information in short-term memory for visually presented material.

How is this construct related to other constructs/ factors? Short-term memory underlies perceptual speed and accuracy. The type of memory measured by this test may not be as deep as working memory

B. **Taxonomy/types/categories of items:** Items vary in the number of symbols in the stimulus set (1, 3, or 5), interval between offset of the stimulus set and onset of the probe (.5 or 2.5 sec), type of symbols used (letters vs. relatively unfamiliar), and correct answer ("in" [the stimulus set] vs "out").

How the test is scored/scaled? Proportion correct and mean of trimmed (of extreme values) decision time are the primary outcome measures.

How are item type and difficulty jointly sampled? The design parameters are crossed.

C. **Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers):** In 256 incumbents in the Project A Field Test, there was a significant correlation between experience with video games and percent correct ($r = .13$) and a non-significant correlation with the latency outcome.

D. **Basis for expecting incremental validity in specific jobs**
In the judgment of Project A's expert panel (Wing, Peterson, and Hoffman, 1984), visual memory will have a median correlation of .21 with 14 tasks of inspecting, troubleshooting, and repairing various kinds of systems. For controlling air traffic and for tasks making use of maps, estimated validities are in the 30s.

2. **Precision of measurement**

A. **Test's consistency across**

Different forms: One form of the test exists now.

Testing occasions: In 474 incumbents, retest reliability was .41 and .66 for the accuracy and latency measures.

Levels of total test score (score conditional SEM): To be determined under the contract that follows up Project A.

B. **For power or speeded test: internal consistency:** .96 and .60 odd-even reliabilities for the latency and percent correct measures

C. **Variation across subgroups in score-conditional SEM:** To be determined under the contract that follows up Project A

D. **If score is model-based, what's the model-based precision:** N.A.

E. **Subgroup differences in item functioning:** Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the latency measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,976	86.94	23.23
Blacks	2,483	88.93	25.16
Hispanic	330	93.25	28.03
Other	348	87.39	24.31
Females	886	87.65	22.09
Males	8,251	87.73	24.24

3. Validity

A. **Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB:** In a factor analysis with pre-enlistment ASVAB and all of the Project A computerized and spatial tests (p. 81), Short-term Memory's proportion correct loaded .3 on a factor named Complex Speed and Accuracy, along with the accuracy and latency scores for Target Identification and Perceptual Speed and Accuracy. Its latency score, however, loaded .37 with a factor otherwise formed of scores from the Simple and Choice Reaction Time tests. These two scores tend to split apart in other factor analyses of subsets of those same tests (pp. 82-5).

B. **Uniqueness re ASVAB. How assessed?** In terms of odd-even reliability, uniqueness of the accuracy and latency outcomes are .55 and .93, in order. In terms of retest reliability, they are .36 and .63, also in order.

C. **Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated?** Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service (1986/87), and then performance tests on their anti-tank gunnery skills

about twelve weeks later. The criterion measure used here was accuracy on the fifth event (event = set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance. The results are shown on p. 100.

Short-term Memory correlated .050 (n.s.) with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. In various GLM analyses with all ASVAB and other Project A tests, Short-term Memory had significant unique variance ($p = .024$). In the stepwise regression, it added .0095 to the R^2, entering second.

D. Subgroup differences in incremental validity or in std err of prediction of the criterion? To be determined under the contract that follows up Project A

E. Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants: It bears repeating that we will do better looking for gains at some level of job analysis finer than the whole job. The tasks for which Project A's expert panel expected Short-term Memory to have useful validities have their counterparts in jobs across the Services.

4. Equating

A. Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated? The design parameters for the items being as objective as they are, automation should be possible. In order to keep the unfamiliar symbolic stimuli from becoming familiar to recruiters and applicants, we may want to enlarge the pool of those symbols.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed? No unusual testing conditions. Samples with sufficient n's in the smaller subgroups

C. Subgroup dependence of equatings: To be determined when we get to the stage of equating.

D. **Practicality of pencil-and-paper versions:** Computerization makes this kind of test possible

5. **Feasibility for operational S/C**

A. **Min./Max. admin times. Times for instrux/test for low and high AFQT examinees:** For the 9,269 cases in the Concurrent Validation, the mean and s.d. of time for the test + instructions were 7m 35s and 1m 27s (p. 104).

B. **Susceptibility to**

Cheating, compromise: As long as the item pool is large enough, probably low

Irrelevant response strategies, coaching: None that we are aware of

Practice: Gains in accuracy and latency were .17 and .15 in 474 incumbents

C. **Floor/ceiling effects? Performance of low/high AFQTs on the test?** Performance by AFQT level will be determined under the contract that follows up Project A. With a percent correct of .88 (s.d. .08) on retesting there may be some squeezing of scores at the top. One s.d. below the retest mean is about .6 sec. How much that could decline is yet to be determined.

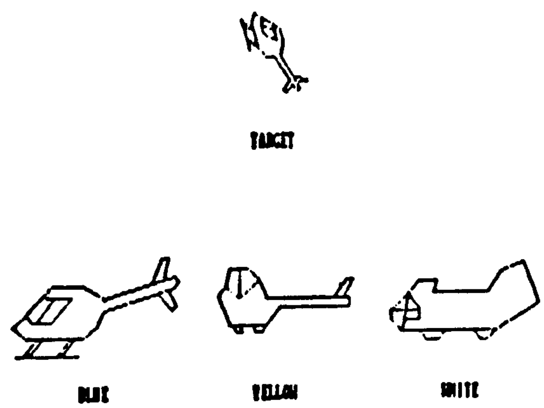
D. **Examinees' reactions: motivation, enjoyment, boredom:** Nothing noteworthy

E. **Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required:** A key arrangement is needed to assure that examinees start their keypresses from a standard hand position.

F. **Need for special eqpt with special maintenance?** None

The case for Target Identification to be in the ECAT Validation

Description: Target Identification is a 48-item test from the computerized battery of Project A that involves pressing a button to pick from three outline figures low on the video screen the one that is just like a target figure at the top of the screen. The three choices may be rotated from the position of the target figure. In order for the stimuli to be presented, the examinee's hands must be in a standard position on the reponse device. Practice items and self-paced instructions are given on the screen. A sample screen is shown below.



1. Theoretical development of construct and measure(s)

A. Definition of construct (process[es]):

Domain: Perceptual speed and accuracy

The ability to make rapid, accurate comparisons between two visual stimuli. Scanning and working memory are involved. For this test, which varies the orientation of the target stimulus, mental rotation is involved.

How is this construct related to other constructs/factors
See below 3.A. Construct Validation

B. Taxonomy/types/categories of items:

Four stimulus dimensions were varied: type of target (e.g., tank, fixed wing aircraft, ...), position of correct response choice (left, middle, right), angle of rotation from the normal upright position, and match between right-left orientation of target and response choices.

How the test is scored/scaled? The primary scores (p.) are Proportion Correct and Mean Decision Time. Decision Time is the time from stimulus onset to onset of hand motion.

How are item type and difficulty jointly sampled: Two items were created for each cell of a table that crosses the design parameters. The parameters all have a significant effect on item difficulty.

C. Seeming differences across subgroups in experiences relevant to any type of item or in offensiveness of content (e.g., experience with computers)

In 256 incumbents in the Project A field test, the accuracy and speed scores on Target Identification correlated .05 - .08 (n.s.) with self reports of experience with video games. Such experience may be related to SES. Offensiveness seems not to be an issue, as the stimuli are planes, helicopters, etc.

D. Basis for expecting incremental validity in specific jobs

- Except that it uses diagrammatic graphics, it bears little resemblance to ASVAB subtests;
- It resembles the military task of identifying real targets quickly and accurately;
- In the judgment of Project A experts (Wing, Peterson, and Hoffman, 1984), perceptual speed and accuracy are likely to predict a variety of aspects of enlisted performance, including: inspecting, troubleshooting, and repairing various types of systems; sending/ receiving radio messages; operating keyboard devices; detecting and identifying targets; firing direct fire weapons; a variety of clerical tasks; translating

or decoding data; and controlling air traffic.

2. Precision of measurement

A. Test's consistency across

Different forms: One form of the test exists now

Testing occasions: For mean decision time, retest reliability in >459 incumbents was .78. Mean % correct had a reliability of .40.

Levels of total test score (score conditional SEM): To be determined under the contract following up Project A.

B. For power or speeded test: internal consistency: Split half reliabilities were .97 and .62 for the time and accuracy scores.

C. Variation across subgroups in score-conditional SEM: To be determined under the contract that follows up Project A.

D. If score is model-based, the model-based precision: N.A.

E. Subgroup differences in item functioning: Fairness analyses will be carried out under the contract that follows up Project A. Means and s.d.'s of the decision time measure by group in the Concurrent Validation were as follows:

Group	n	$\bar{x}$	s.d.
Whites	5,939	180.19	55.05
Blacks	2,479	222.45	68.64
Hispanics	324	213.08	68.51
Other	351	201.38	71.61
Females	878	223.33	66.23
Males	8,215	190.53	61.98

3. Validity

A. Construct validity: correlations with other tests of similar things. Intercorr. with ASVAB: Mean decision time loads most heavily on the factor of perceptual speed/accuracy, but has moderate secondary loadings with spatial and psychomotor tests (pp 81-5). Among the ASVAB subtests (p. 87), the decision time score correlates highest with MC (.33), GS (.30), and AS (.25).

B. Uniqueness re ASVAB. How assessed? .83 in terms of odd-even reliability; .64 relative to retest reliability.

C. Incremental validity re ASVAB? Improvement in SD of the predicted criterion? How many other tests were considered when this one was validated? Incremental validities of all computerized and spatial predictors from Project A were examined in GLM analyses by Silva (1989, unpublished). The sample was 311 trainees who had taken ASVAB as applicants, the Project A tests during their first three days of Service, and then performance tests of their anti-tank gunnery skills about twelve weeks later. The criterion measure was an accuracy score on the fifth event (i.e., set of 10 trials), which was the first event on which a gunner could attain qualification. Correlations of all Project A tests, AFQT, and ASVAB's Combat Composite with the criterion were first examined. Variance in the criterion accounted for by two sets of tests - all Project A and all ASVAB - was computed separately and then same analysis was run on the Project A tests with the criterion variance associated with ASVAB removed. Finally, the Project A tests were entered into a stepwise linear regression on the residual (unrelated to ASVAB) variance. Results are shown on p. 100.

The decision time score for Target Identification correlated .174 with the criterion, compared with .179 for the Combat Composite (CS + AR + MC + AS) and .145 with AFQT. The multiple R^2 for all ASVAB subtests was .041, for all Project A was .166, and for both sets together .177. Project A tests accounted for .126 of the residual variance.

The unique variance accounted for by Target Identification had a p between .05 and .10 when all ASVAB and other Project A tests were included in the multiple regression and in analyses of Project A tests alone. In a stepwise analysis of Project A tests predicting the non-ASVAB related criterion variance, Target ID is the third of five tests to enter ($p = .0403$), adding .0125 to the R^2.

D. Subgroup differences in incremental validity or in std err of prediction of the criterion? To be determined under the contract that follows up Project A.

E. Job families and patterns of validities for them. Similarity of job families in the other Services. Expected utility in terms of numbers, attrition, training costs, attractiveness to applicants: See above pp. 38-9. Target identification is itself a job task in air defense, tank and anti-tank gunnery, and fire support. As an indicator of perceptual speed and accuracy, this test should predict performance in radio operating, typing, translating, decoding, and controlling air traffic. These are all pan-Service activities.

4. Equating

A. Problems in developing equivalent forms (e.g., How large is the potential item pool? Would pretesting and vetting of new items be needed? Could item devel be automated?) The potential item pool is very large when you multiply types of targets (truck, helicopter, tank, etc.) by orientation by composition of the set of response choices by position of the correct choice. New items would need to be pretested at least for debugging. Whether the existing item pool is large enough to permit estimation of item parameters for all other potential items needs to be determined.

B. Types of samples needed/acceptable for initial equating? Testing conditions needed? Adequate numbers of minorities and low AFQT examinees will be hardest to get. Existing data from ASVAB R&D should show whether recruits respond like applicants.

C. Subgroup dependence of equatings: To be determined when equating becomes an issue for this test.

D. Practicality of pencil-and-paper versions
In principle this test could be given on paper, but that would result in loss of the best response measure: mean decision time.

5. Feasibility for operational S/C:

A. Min./Max. admin times. Times for instrux/test for low and high AFQT examinees: Total test time averages 4m 33s, s.d. 1m 15s. Correlation of time to read the directions with AFQT is around .30.

B. Susceptibility to:

Cheating, compromise: Given the large number of potential items, the hazard seems low.

Irrelevant response strategies: Validity could be undermined by examinees' sacrificing accuracy for speed. To counter that tactic, a composite predictor could be developed from latency adjusted for accuracy.

Coaching: Doubtful

Practice: The gain on retesting 473 incumbents was .32 s.d. The relative contribution to that gain of familiarity with items and test-wiseness needs to be examined.

C. Floor/ceiling effects? Performance of low/high AFQTs

on the test? Maximum mean decision time is great enough not to be limited. On retesting, the mean time was about 1.3 sec longer than mean decision time for the retest of choice reaction time.

D. Examinees' reactions: motivation, enjoyment, boredom: Nothing noteworthy

E. Limitations of CAT-ASVAB hardware for delivering the test. Additional hardware required: This test uses the four home buttons on the reponse pedestal, to assure a standard hand position for responding, plus three response buttons.

F. Need for special eqpt with special maintenance? No other

Factor Analysis[1] of ASVAB Subtests, Cognitive Paper-and-Pencil Tests, and Computer-Administered Measures

	FACTOR I General Cognition	FACTOR II Spatial	FACTOR III Number	FACTOR IV Psychomotor	FACTOR V Complex Speed & Accuracy	FACTOR VI Simple Speed & Accuracy	h^2
ASVAB: General Science	.77	.17	-.09	-.10	.03	-.09	.66
ASVAB: Word Knowledge	.74	.06	-.11	-.02	.09	-.12	.60
ASVAB: Electronics Information	.64	.16	.02	-.17	.01	.02	.47
ASVAB: Mechanical Comprehension	.62	.43	00	-.25	.02	.03	.64
ASVAB: Paragraph Comprehension	.62	.07	-.17	-.01	.08	-.12	.44
ASVAB: Auto/Shop	.62	.19	.10	-.21	00	.04	.48
ASVAB: Arithmetic Reasoning	.52	.35	-.44	-.04	.13	.03	.62
ASVAB: Mathematics Knowledge	.47	.34	-.46	-.01	.13	00	.58
Assembling Objects	.22	.64	-.09	-.17	.12	-.09	.53
Maze Test	.12	.57	-.15	-.31	-.01	-.08	.48
Reasoning Test	.32	.55	-.22	-.13	.13	-.11	.50
Object Rotation	.14	.51	-.11	-.24	-.03	-.05	.35
Map Test	.46	.50	-.21	-.16	.11	-.05	.55
Orientation Test	.34	.50	-.10	-.16	.06	-.02	.41
Number Memory: Operations Time Pooled Mean	-.22	-.02	.69	.07	.21	.02	.58
Number Memory: Final Response Time Mean	-.15	-.16	.57	.17	.17	.19	.48
Number Memory: Input Response Time Mean	-.14	-.01	.51	.16	.05	.04	.31
Number Memory: Proportion Correct	.15	.11	-.39	-.05	.31	-.02	.30
ASVAB: Coding Speed	-.04	.20	-.45	00	.08	-.14	.28
ASVAB: Number Operations	-.15	.09	-.60	.01	.06	-.06	.41
Target Tracking 1: Mean Log (Distance + 1)	-.17	-.15	.03	.80	-.05	.09	.71
Target Tracking 2: Mean Log (Distance + 1)	-.22	-.17	.05	.77	-.02	.07	.68
Target Shoot: Mean Log (Distance + 1)	-.06	-.04	.01	.50	-.17	.08	.30
Cannon Shoot: Mean Absolute Time Discrepancy	-.11	-.17	.11	.45	-.02	.09	.27
Target Shoot: Mean Time to Fire	-.12	-.18	.11	.34	.22	.06	.23
Pooled Mean Movement Time	.04	-.14	.10	.26	.08	.07	.11
Perceptual Speed & Accuracy: Decision Time Mean	.01	-.11	.18	.06	.68	.22	.57
Perceptual Speed & Accuracy: Proportion Correct	.04	.09	-.08	-.05	.66	-.03	.46
Target Identification: Proportion Correct	.07	.12	00	-.06	.52	00	.29
Target Identification: Decision Time Mean	-.19	-.34	.11	.29	.48	.22	.55
Short Term Memory: Proportion Correct	.09	.17	-.12	-.09	.30	-.17	.19
Choice Reaction Time: Proportion Correct	.12	.06	-.01	-.02	.18	-.13	.07
Choice Reaction Time: Decision Time Mean	.01	-.03	.19	.12	.14	.55	.37
Simple Reaction Time: Decision Time Mean	00	-.02	.04	.09	.01	.45	.21
Short Term Memory: Decision Time Mean	.06	-.13	.21	.09	.35	.37	.34
Simple Reaction Time: Proportion Correct	.12	.04	00	-.01	.09	-.23	.08
h^2	4.08	2.77	2.63	2.51	2.06	1.03	15.09

[1]Princip. factor analysis, communality estimate = squared multiple correlation, varimax rotation.

Note: N = 7082

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Factor Analysis[1] of ASVAB Subtests and Computer-Administered Measures

	FACTOR I General Cognition	FACTOR II Number	FACTOR III Psychomotor	FACTOR IV Complex Speed & Accuracy	FACTOR V Simple Speed & Accuracy	h^2
ASVAB: General Science	.79	.11	-.10	.04	-.09	.65
ASVAB: Word Knowledge	.73	.11	-.01	.10	-.11	.57
ASVAB: Mechanical Comprehension	.69	.07	-.32	.03	00	.59
ASVAB: Electronics Information	.66	00	-.18	.02	.01	.48
ASVAB: Auto/Shop	.65	-.06	-.23	00	.03	.48
ASVAB: Paragraph Comprehension	.61	.18	00	.08	-.11	.43
ASVAB: Arithmetic Reasoning	.57	.49	-.11	.13	.02	.60
ASVAB: Number Operations	-.14	.62	00	.06	-.07	.41
ASVAB: Mathematics Knowledge	.51	.52	-.08	.13	00	.56
ASVAB: Coding Speed	00	.49	-.03	.08	-.14	.27
Number Memory: Proportion Correct	.16	.41	-.07	.32	-.02	.30
Number Memory: Input Response Time Mean	-.12	-.48	.13	.05	.04	.28
Number Memory: Final Response Time Mean	-.17	-.58	.20	.16	.19	.48
Number Memory: Operations Time Pooled Mean	-.20	-.66	.06	.21	.01	.54
Target Tracking 1: Mean Log (Distance + 1)	-.18	-.02	.81	-.07	.08	.71
Target Tracking 2: Mean Log (Distance + 1)	-.23	-.04	.78	-.04	.06	.68
Target Shoot: Mean Log (Distance + 1)	-.05	00	.50	-.17	.06	.29
Cannon Shoot: Mean Absolute Time Discrepancy	-.13	-.11	.47	-.03	.09	.27
Target Shoot: Mean Time to Fire	-.15	-.11	.37	.21	.07	.23
Pooled Mean Movement Time	.01	-.11	.29	.08	.07	.11
Perceptual Speed & Accuracy: Decision Time Mean	00	-.19	.11	.69	.22	.57
Perceptual Speed & Accuracy: Proportion Correct	.05	.11	-.05	.67	-.04	.46
Target Identification: Proportion Correct	.10	.01	-.06	.51	00	.28
Target Identification: Decision Time Mean	-.26	-.15	.36	.47	.23	.50
Short Term Memory: Proportion Correct	.12	.16	-.12	.30	-.17	.18
Choice Reaction Time: Proportion Correct	.13	.02	-.02	.19	-.13	.07
Choice Reaction Time: Decision Time Mean	.02	-.18	.13	.14	.55	.37
Simple Reaction Time: Decision Time Mean	00	-.03	.10	00	.45	.22
Short Term Memory: Decision Time Mean	.04	-.22	.14	.35	.37	.34
Simple Reaction Time: Proportion Correct	.12	.01	-.01	.09	-.23	.08
h^2	3.86	2.64	2.48	2.00	1.00	11.98

[1] Principle factor analysis, communality estimate = squared multiple correlation, varimax rotation.

Note: N = 7119

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Factor Loading[1] Matrix

	Spatial Ability	Psychomotor Ability	Complex Perceptual Speed & Accuracy	Numerical Speed and Accuracy	Simple Reaction Speed	Simple Reaction Accuracy
Assembling Objects	.77	-.11	.12	-.01	-.06	.01
Reasoning Test	.72	-.09	.14	-.18	-.05	.08
Map Test	.70	-.14	.13	-.19	-.01	.05
Orientation Test	.69	-.12	.05	-.06	.02	.03
Maze Test	.67	-.27	-.03	-.09	-.07	-.02
Object Rotation	.63	-.16	-.05	-.07	-.03	.00
Target Tracking 1						
Mean Log (Distance + 1)	-.25	.84	.00	.05	.05	-.05
Target Tracking 2						
Mean Log (Distance + 1)	-.28	.82	.03	.06	.03	-.04
Target Shoot						
Mean Log (Distance + 1)	-.03	.68	-.18	.00	.06	.01
Mean Time to Fire	-.27	.36	.31	.15	.00	-.07
Cannon Shoot						
Absolute Time Discrepancy	-.20	.57	-.05	.13	.09	.05
Perceptual Speed & Accuracy						
Proportion Correct	.12	-.07	.74	-.02	.03	.09
Trimmed Decision Time Mean	-.13	.02	.71	.21	.27	.05
Target Identification						
Proportion Correct	.15	-.08	.64	.08	.03	.03
Trimmed Decision Time Mean	.45	.26	.53	.16	.21	-.05
Number Memory						
Proportion Correct	.17	-.04	.47	-.43	-.08	-.04
Pooled Mean Operation Time	-.13	.02	.15	.83	.04	-.04
Input Time Mean	-.08	.13	.00	.76	.00	-.07
Final Response Time Mean	-.26	.11	.07	.68	.25	.08
Short Term Memory						
Proportion Correct	.23	-.08	.41	-.12	-.17	.12
Trimmed Decision Time Mean	-.13	.03	.32	.23	.53	.12
Choice Reaction Time						
Trimmed Decision Time Mean	-.03	.10	.08	.14	.76	-.01
Proportion Correct	.10	-.04	.12	-.07	.06	.72
Simple Reaction Time						
Trimmed Decision Time Mean	.00	.08	-.06	-.06	.74	-.11
Proportion Correct	.06	-.02	.02	.01	-.21	.71
Pooled Mean Movement Time	-.13	.29	.02	.06	.22	.33

[1] Principal component analysis, varimax rotation.

Note: N = 8372

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Table 10

Factor Loading Matrix Defining ACB Predictor Composites Based on Principal Components Analysis with Varimax Rotation

	Spatial Abilities	Tracking/ Psychomotor	Perceptual Speed and Accuracy	Numerical Operations	Reaction Time	Movement Time
Reasoning Test	.65	-.03	.29	-.06	-.24	.12
Object Rotation	.59	-.13	-.08	-.08	.08	-.13
Orientation Test	.75	-.15	.12	-.06	-.20	-.08
Maze Test	.22	-.35	.13	.04	-.21	.04
Map Test	.49	-.30	-.10	-.23	-.12	.01
Simple Reaction Time						
Percent Correct	.15	.09	.19	.08	-.38	-.02
Decision Time Mean	-.02	.19	.03	.31	.56	.12
Choice Reaction Time						
Percent Time Mean	.28	.21	.21	.07	-.13	.26
Decision Time Mean	-.02	.25	.28	.39	.54	.15
Memory						
Percent Correct	.13	-.05	.06	.07	-.56	.25
Decision Time Mean	.22	-.12	.40	.35	.23	.39
Tracking 1	-.23	.86	.10	.14	-.02	-.06
Perceptual Speed						
Percent Correct	.19	-.10	.57	-.04	-.06	-.13
Decision Time Mean	-.28	-.07	.62	.21	.06	.17
Tracking 2	-.25	.75	.05	.11	.08	.01
Number Memory						
Percent Correct	.28	-.10	.63	-.09	-.16	.03
Final Response Time Mean	-.19	.07	.01	.53	.18	.52
Pooled Operations Mean	-.08	-.14	.09	.73	.06	.09
Input Response Time Mean	-.15	.13	.04	.70	-.05	-.09
Cannon Shoot						
Mean Absolute Time Discrepancy	-.06	.40	-.11	-.05	.13	.08
Target Identification						
Percent Correct	.05	-.01	.46	.06	-.06	.03
Decision Time Mean	-.34	.17	.56	.20	.20	-.02
Target Shoot						
Mean Time To Fire	-.45	.46	.17	.14	.08	.22
Mean Log (Distance + 1)	.16	.62	-.13	-.16	.05	-.01
Pooled Mean Movement Time	-.11	.07	-.02	-.04	-.09	.74

Source: Smith & Graham, 1987

Concurrent Validity Data Analysis

Factor Analysis[1] of Computer-Administered Measures

	FACTOR I Psychomotor	FACTOR II Perceptual Speed	FACTOR III Perceptual Accuracy	h^2
Target Tracking 1: Mean Log (Distance + 1)	.83	.08	-.16	.72
Target Tracking 2: Mean Log (Distance + 1)	.81	.11	-.13	.68
Target Shoot: Mean Log (Distance + 1)	.50	-.01	-.20	.30
Cannon Shoot: Mean Absolute Time Discrepancy	.47	.17	-.14	.27
Target Shoot: Mean Time to Fire	.38	.25	.09	.22
Pooled Mean Movement Time	.27	.16	.01	.09
Number Memory: Operations Time Pooled Mean	.01	.70	-.08	.50
Number Memory: Final Response Time Mean	.16	.68	-.10	.50
Number Memory: Input Response Time Mean	.07	.56	-.21	.36
Short Term Memory: Decision Time Mean	.16	.45	.24	.29
Target Identification: Decision Time Mean	.43	.43	.31	.46
Choice Reaction Time: Decision Time Mean	.19	.38	.08	.19
Simple Reaction Time: Decision Time Mean	.15	.19	00	.06
Perceptual Speed and Accuracy: Proportion Correct	-.01	.12	.67	.46
Perceptual Speed and Accuracy: Decision Time Mean	.16	.46	.58	.58
Target Identification: Proportion Correct	-.05	.16	.50	.28
Number Memory: Proportion Correct	-.05	-.24	.43	.25
Short Term Memory: Proportion Correct	-.13	-.12	.36	.16
Choice Reaction Time: Proportion Correct	-.05	-.04	.23	.06
Simple Reaction Time: Proportion Correct	-.07	-.08	.11	.03
h^2	2.38	2.32	1.76	6.46

[1]Principle factor analysis, communality estimate = squared multiple correlation, varimax rotation.

Note: N = 8521

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Correlations Between Cognitive Paper and Pencil Measures and ASVAB Subtests
(N = 7939)

ASVAB Subtest	Assembling Objects	Object Rotation	Maze Test	Orientation Test	Map Test	Reasoning Test
Word Knowledge	.25	.16	.17	.31	.42	.35
Paragraph Comprehension	.22	.16	.19	.28	.37	.32
Coding Speed	.17	.15	.23	.11	.19	.19
Number Operations	.06	.09	.14	.03	.10	.00
Arithmetic Reasoning	.40	.30	.33	.43	.54	.51
Mathematics Knowledge	.39	.25	.30	.41	.52	.47
Mechanical Comprehension	.45	.38	.40	.48	.55	.46
General Science	.32	.25	.27	.37	.49	.40
Auto/Shop Information	.27	.27	.26	.33	.40	.27
Electronics Information	.26	.22	.23	.31	.38	.28
AFQT[a]	.37	.27	.31	.41	.31	.49

[a]AFQT is a composite of scores from ASVAB subtests Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning and Numerical Operations.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Correlations Between Computer-Administered Measures and ASVAB Subtests

	NO	CS	GS	AR	WK	PS	AS	MK	MC	EI
Simple Reaction Time: Decision Time Mean	.05	.05	.06	.07	.04	.04	.05	.06	.07	.04
Simple Reaction Time: Proportion Correct	.01	.04	.11	.11	.11	.11	.10	.08	.10	.08
Choice Reaction Time: Decision Time Mean	.16	.16	.07	.08	.05	.07	-.01	.10	.05	.02
Choice Reaction Time: Proportion Correct	.01	.05	.12	.10	.11	.11	.10	.07	.10	.09
Choice Reaction Time (Decision)-Simple Reaction Time (Decision)	.06	.06	-.01	-.02	.00	.01	-.05	.00	-.03	-.02
Perceptual Speed & Accuracy: Decision Time Mean	.12	.12	.04	.02	.00	.03	-.01	.04	.02	.00
Perceptual Speed & Accuracy: Proportion Correct	.09	.14	.10	.16	.13	.13	.06	.16	.08	.07
Short Term Memory: Decision Time Mean	.13	.13	.03	.06	-.01	.01	.00	.08	.03	.01
Short Term Memory: Proportion Correct	.08	.13	.15	.19	.16	.15	.09	.19	.15	.10
Target Identification: Decision Time Mean	.06	.10	.30	.20	.19	.18	.25	.19	.33	.21
Target Identification: Proportion Correct	.02	.06	.13	.13	.13	.10	.10	.14	.13	.10
Number Memory: Input Response Time Mean	.20	.19	.17	.25	.19	.18	.09	.22	.15	.10
Number Memory: Operations Time Pooled Mean	.33	.21	.24	.41	.21	.22	.13	.39	.17	.14
Number Memory: Final Response Time Mean	.30	.26	.24	.37	.21	.21	.16	.35	.21	.17
Number Memory: Proportion Correct	.27	.15	.18	.40	.18	.18	.12	.35	.17	.13
Pooled Mean Movement Time	.07	.09	.06	.08	.00	.03	.08	.09	.11	.07
Target Tracking 1: Mean Log (Distance + 1)	.02	.07	.28	.25	.20	.18	.30	.22	.40	.27
Target Tracking 2: Mean Log (Distance + 1)	.02	.08	.31	.27	.23	.21	.33	.24	.41	.30
Target Shoot: Mean Log (Distance + 1)	.01	.06	.15	.12	.11	.09	.12	.12	.20	.14
Target Shoot: Mean Time to Fire	.06	.08	.17	.15	.12	.11	.19	.12	.23	.15
Cannon Shoot: Mean Absolute Time Discrepancy	.05	.07	.19	.22	.15	.13	.22	.19	.30	.19

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Note: N range = 7534-7863. Distance and time score correlations have been reflected.

Con current Validity Data Analysis

Factor Loadings for Paper and Pencil Cognitive Tests-Score Correct
Two-Factor Solution with Orthogonal Varimax Rotation

Test	Factor 1 Loading	Factor 2 Loading	h^2
Map Test	60	37	.50
Reasoning Test	59	40	.50
Orientation Test	56	34	.43
Assembling Objects	54	47	.51
Maze Test	38	57	.48
Object Rotation	32	52	.38
h^2	1.56	1.24	

Cognitive Paper and Pencil Measures
Score Correct Intercorrelations

Test	Object Rotation	Maze Test	Orientation Test	Map Test	Reasoning Test
Assembling Objects	.41	.51	.46	.50	.56
Object Rotation		.50	.37	.39	.38
Maze Test			.40	.44	.45
Orientation Test				.53	.48
Map Test					.52

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Factor Loading[1] Matrix for Cognitive Test Scores

Defining Alternative Composite

	Factor A	Factor B	Factor C
Assembling Objects	$\boxed{.81}$	.23	.25
Reasoning Test	$\boxed{.76}$	.37	.13
Orientation Test	.20	$\boxed{.87}$	.20
Map Test	.42	$\boxed{.69}$	.20
Object Rotation	.12	.25	$\boxed{.89}$
Maze Test	.52	.12	$\boxed{.66}$

1 Principal components analysis, varimax rotation.

Note: N = 9301

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Correlations of Thirteen Computer-Administered Psychomotor Test Scores

Score	MRD	ML(D+1)	MRD	ML(D+1)	MRD	ML(D+1)	MTF	H/MP	MMD	ML(D+1)	MTD	MATS	H/MP
Target Tracking 1 - Mean Raw Distance	1.00												
Target Tracking 1 - Mean Log (Distance + 1)	.95	1.00											
Target Tracking 2 - Mean Raw Distance	.71	.77	1.00										
Target Tracking 2 - Mean Log (Distance + 1)	.68	.77	.97	1.00									
Target Shoot - Mean Raw Distance	.34	.34	.32	.30	1.00								
Target Shoot - Mean Log (Distance + 2)	.44	.48	.44	.44	.87	1.00							
Target Shoot - Mean Time to Fire	.28	.36	.33	.36	-.13	-.06	1.00						
Target Shoot - Hit/Miss[1] Proportion	.47	.53	.48	.49	.59	.84	-.07	1.00					
Cannon Shoot - Mean Minimum Distance	.39	.43	.42	.42	.25	.32	.18	.33	1.00				
Cannon Shoot - Mean Log (Distance + 2)	.34	.38	.37	.38	.21	.27	.17	.28	.93	1.00			
Cannon Shoot - Mean Absolute Time Discrepancy	.40	.45	.43	.43	.26	.33	.18	.33	.96	.90	.19	1.00	
Cannon Shoot - Hit/Miss[1] Proportion	.29	.34	.33	.34	.17	.23	.16	.24	.82	.87	.31	.81	1.00

Note: Ns for these correlations range from 8878 to 9251.

[1] Hit/Miss Proportion scores' correlations with distance measures have been reflected.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis
Means, Standard Deviations and Reliability Estimates
for Computerized Psychomotor Tests

	N	Mean	Standard Deviation	Odd-Even Reliability	Test-Retest Reliability[1]
Target Tracking 1:					
Mean Raw Distance	9251	24.43	15.57	.97	.61
Mean Log (Distance +1)	9251	2.98	.49	.98	.74
Target Tracking 2:					
Mean Raw Distance	9239	49.25	24.25	.97	.82
Mean Log (Distance + 1)	9239	3.70	.51	.98	.85
Target Shoot:					
Mean Raw Distance	8892	10.23	5.78	.92	.19
Mean Log (Distance + 1)	8892	2.17	.24	.74	.37
Mean Time to Fire	8892	235.39	47.78	.85	.58
Hit/Miss Proportion	8892	.53	.13	.52	.44
Cannon Shoot:					
Mean Minimum Distance	9235	8.41	1.85	.69	.53
Mean Log (Minimum Distance + 1)	9235	1.93	.18	.58	.46
Mean Time Discrepancy (+ and -)	9234	5.55	14.25	.76	.48
Mean Absolute Time Discrepancy	9234	43.94	9.57	.65	.52
Hit/Miss Proportion	9235	.55	.10	.49	.43

Note: Raw distance measures are in units = 1/200th vertical screen distance, and time-to-fire and time-discrepancy measures are in hundredths of seconds. Logs are natural logs.

[1]Test-Retest reliability estimates are based on sample sizes of 468 to 487. Retest interval was 4 to 6 weeks.

Concurrent Validity Sample
Means, Standard Deviations and Reliability Estimates for Computerized Cognitive Perceptual Tests

	N	Mean[1]	Standard Deviation	Split-Half Reliability	Test-Retest Reliability[2]
Simple Reaction Time:					
Decision Time Mean	9255	31.84	14.82	.88	.23
Decision Time-Ind-SD	9255	8.43	17.86	.84	.11
Movement Time Mean	9255	26.51	12.04	.56	.24
Movement Time-Ind-SD	9255	11.22	28.60	.31	.01
Log Movement Time Mean	9255	3.16	0.26	.92	.59
Log Movement Time Ind-SD	9255	0.22	0.19	.43	.07
Proportion Correct	9255	0.98	0.04	.46	.02
Choice Reaction Time:					
Decision Time Mean	9269	40.93	9.77	.97	.69
Decision Time-Ind-SD	9269	8.68	6.21	.89	.41
Movement Time Mean	9269	27.04	7.52	.61	.46
Movement Time-Ind-SD	9269	13.92	21.48	.24	.17
Log Movement Time Mean	9269	3.19	0.22	.93	.67
Log Movement Time Ind-SD	9269	0.30	0.14	.48	.28
Proportion Correct	9269	0.98	0.03	.57	.23
Choice DT Mean - Simple DT Mean	9250	9.09	14.44	.85	
Short Term Memory:					
Decision Time Mean	9149	87.72	24.03	.96	.66
Decision Time-Ind-SD	9149	27.77	13.57	.83	.38
Movement Time Mean	9149	40.97	13.57	.92	.55
Movement Time-Ind-SD	9149	16.93	13.73	.36	.12
Log Movement Time Mean	9149	3.60	0.31	.97	.63
Log Movement Time Ind-SD	9149	0.30	0.10	.52	.30
Proportion Correct	9149	0.89	0.08	.60	.41
Perceptual Speed & Accuracy:					
Decision Time Mean	9244	236.91	63.38	.94	.63
Decision Time-Ind-SD	9244	134.81	46.89	.89	.56
Movement Time Mean	9244	35.63	11.57	.87	.61
Movement Time-Ind-SD	9244	17.46	16.26	.27	.15
Log Movement Time Mean	9244	3.45	0.29	.96	.66
Log Movement Time Ind-SD	9244	0.32	0.12	.47	.24
Proportion Correct	9244	0.87	0.08	.65	.51

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Means, Standard Deviations and Reliability Estimates for Computerized Cognitive Perceptual Tests (Continued)

	N	Mean[1]	Standard Deviation	Split-Half Reliability	Test-Retest Reliability[2]
Target Identification:					
Decision Time Mean	9105	193.65	63.13	.97	.78
Decision Time-Ind-SD	9105	67.01	31.77	.86	.62
Movement Time Mean	9105	38.05	11.39	.87	.51
Movement Time-Ind-SD	9105	18.01	16.31	.32	.14
Log Movement Time Mean	9105	3.53	0.26	.96	.62
Log Movement Time Ind-SD	9105	0.30	0.12	.44	.25
Proportion Correct	9105	0.91	0.07	.62	.40
Number Memory:					
Final Response Time Mean	9099	160.70	42.63	.88	.62
Final Response Time Ind-SD	9099	65.99	37.42	.78	.49
Input Response Time Mean	9099	142.84	55.24	.95	.47
Input Response Time Ind-SD	9099	51.53	34.79	.85	.30
Add Operations Mean	9099	186.33	73.23	.84	.70
Add Operations Ind-SD	9099	96.40	59.92	.58	.59
Subtract Operations Mean	9099	226.90	86.46	.80	.66
Subtract Operations Ind-SD	9099	131.88	74.23	.52	.40
Multiplication Operations Mean	9099	257.68	95.45	.75	.64
Multiplication Operations Ind-SD	9099	140.61	72.71	.48	.45
Division Operations Mean	9099	261.50	102.75	.77	.64
Division Operations Ind-SD	9099	144.49	77.97	.50	.50
Operations Pooled Mean	9099	233.10	79.71	.93	.73
Operations Pooled Ind-SD	9099	137.20	55.90	.74	.67
Proportion Correct	9099	.90	.09	.59	.53

[1] Times are given in hundredths of seconds. Logs are natural logs.

[2] N=460-479 for test-retest correlations. Retest interval was 4 to 6 weeks.

[3] Coefficient Alpha reliability estimates.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Cognitive Paper-and-Pencil Measures: Odd-Even Reliability Coefficients, Squared Multiple Regression Coefficients (Against All ASVAB Subtests)

Test	N	Odd-Even Reliability	Squared Multiple R[1]	Uniqueness Estimate
Assembling Objects	9343	.91	.26	.65
Object Rotation	9345	.98	.18	.80
Maze Test	9344	.96	.22	.74
Orientation Test	9341	.89	.29	.60
Map Test	9343	.90	.44	.46
Reasoning Test	9332	.87	.34	.53

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

---

[1] Ns for these calculations are 7329.

Concurrent Validity Data Analysis

Odd-Even Reliability Coefficients, Squared Multiple Regression Coefficients (Against All ASVAB Subtests), and Uniqueness Estimates for Thirteen Computer-Administered Psychomotor Test Scores

Score	N	Odd-Even Reliability	Squared[1] Multiple R	Uniqueness[2] Estimate
Target Tracking 1:				
Mean Raw Distance	9250	.97	.12	.85
Mean Log (Distance +1)	9250	.98	.16	.82
Target Tracking 2:				
Mean Raw Distance	9238	.97	.17	.80
Mean Log (Distance + 1)	9238	.98	.19	.79
Target Shoot:				
Mean Raw Distance	8892	.92	.02	.90
Mean Log (Distance + 1)	8892	.74	.04	.70
Mean Time to Fire	8892	.85	.07	.78
Hit/Miss Proportion	8892	.52	.07	.45
Cannon Shoot:				
Mean Minimum Distance	9234	.69	.09	.60
Mean Log (Minimum Distance + 1)	9234	.58	.07	.51
Mean Time Discrepancy (+ and −)	9234	.76	.01	.75
Mean Absolute Time Discrepancy	9234	.65	.09	.56
Hit/Miss Proportion	9234	.49	.06	.43

[1] N = 7516 for these calculations.

[2] Uniqueness = Reliability minus Squared Multiple R

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Odd-Even Reliability Coefficients, Squared Multiple Regression Coefficients (Against All ASVAB Subtests), and Uniqueness Estimates for Sixteen Computer Administered Cognitive/Perceptual Test Scores

Score	N	Odd-Even Reliability	Squared[1] Multiple R	Uniqueness Estimate
Simple Reaction Time:				
Trimmed Decision Time Mean	9254	.88	.01	.87
Proportion Correct	9254	.46	.02	.44
Choice Reaction Time:				
Trimmed Decision Time Mean	9268	.97	.04	.93
Proportion Correct	9268	.57	.02	.55
Choice (DT) - Simple (DT)	9249	.85	.01	.84
Short Term Memory:				
Trimmed Decision Time Mean	9148	.96	.03	.93
Proportion Correct	9148	.60	.05	.55
Perceptual Speed & Accuracy:				
Trimmed Decision Time Mean	9243	.94	.02	.92
Proportion Correct	9243	.65	.04	.61
Target Identification:				
Trimmed Decision Time Mean	9104	.97	.14	.83
Proportion Correct	9104	.62	.03	.59
Number Memory:				
Final Response Time Mean	9099	.88	.21	.67
Input Time Mean	9099	.95[2]	.10	.85
Pooled Operation Time Mean	9099	.91[2]	.25	.66
Proportion Correct	9099	.59	.20	.39
SRT-CRT-STM-PSA-TID:				
Pooled Movement Time Mean	8962	.74[2]	.03	.71

[1] Ns for these calculations range from 7715 to 7860.

[2] These are Coefficient Alpha reliabilities.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

Cognitive Paper and Pencil Measures: Gain Scores for Test-Retest

		Time 1		Time 2		
	N	Mean	SD	Mean	SD	Gain[a]
Assembling Objects	502	23.41	6.60	23.96	6.70	.08
Object Rotation	499	61.83	19.03	68.02	17.88	.33
Maze Test	502	16.54	4.54	17.46	4.90	.20
Orientation Test	500	11.45	6.09	13.11	6.63	.27
Map Test	502	7.77	5.43	8.21	5.86	.08
Reasoning Test	500	19.16	5.48	20.64	5.53	.27

[a]Gain = $(M_2 - M_1)/SD_1$

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

N's, Means, Standard Deviations, and Gain Scores
For Thirteen Computer-Administered Psychomotor Test Scores

Test Score	Time 1 N	Time 1 Mean	Time 1 Standard Deviation	Time 2 N	Time 2 Mean	Time 2 Standard Deviation	Gain Score[1,2]
Target Tracking 1:							
Mean Raw Distance	487	24.15	15.97	487	22.69	17.40	0.09
Mean Log (Distance +1)	487	2.98	0.48	487	2.85	0.54	0.27
Target Tracking 2:							
Mean Raw Distance	487	48.48	25.59	487	45.89	29.20	0.10
Mean Log (Distance + 1)	487	3.68	0.53	487	3.55	0.63	0.24
Target Shoot:							
Mean Raw Distance	468	9.78	2.67	470	10.71	8.82	--0.35
Mean Log (Distance + 1)	468	2.16	0.18	470	2.16	0.32	0.00
Mean Time to Fire	468	236.58	47.99	470	238.62	48.96	--0.04
Hit/Miss Proportion	468	0.53	0.11	470	0.54	0.15	0.07
Cannon Shoot:							
Mean Minimum Distance	487	8.50	1.99	484	7.76	1.86	0.37
Mean Log (Minimum Distance + 1)	487	1.93	0.19	484	1.86	0.19	0.40
Mean Time Discrepancy (+ and -)	487	5.76	15.76	484	2.27	13.55	0.22
Mean Absolute Time Discrepancy	487	44.33	10.12	484	41.34	9.74	0.30
Hit/Miss Proportion	487	0.55	0.10	484	0.58	0.11	0.30

Note: Raw distance measures are in units = 1/200th vertical screen distance, and time-to-fire and time-discrepancy measures are in hundredths of seconds.

[1] Gain Score = (Time 2 Mean -- Time 1 Mean)/Time 1 Standard Deviation

[2] Gain Scores have been appropriately reflected so that positive gain scores signify improvement and negative gain scores signify decrement in performance.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

Concurrent Validity Data Analysis

N's, Means, Standard Deviations, and Gain Scores For Sixteen Computer-Administered Cognitive Test Scores

Score	Time 1 N	Time 1 Mean	Time 1 Standard Deviation	Time 2 N	Time 2 Mean	Time 2 Standard Deviation	Gain Score[1,2]
Simple Reaction Time:							
Decision Time Mean	479	30.63	8.33	486	31.96	9.39	-.16
Proportion Correct	479	.98	.05	486	1.00	.02	.29
Choice Reaction Time:							
Decision Time Mean	479	40.91	8.04	487	42.58	8.51	-.21
Proportion Correct	479	.98	.04	487	.99	.02	.26
Choice DT Mean - Simple DT Mean	479	10.28	9.36	487	10.63	9.63	-.04
Short Term Memory:							
Decision Time Mean	474	85.37	24.63	483	81.63	21.92	.15
Proportion Correct	474	.88	.08	483	.89	.08	.17
Perceptual Speed & Accuracy:							
Decision Time Mean	478	238.19	66.77	484	232.63	67.48	.08
Proportion Correct	478	.87	.09	484	.88	.10	.11
Target Identification							
Decision Time Mean	473	193.48	67.09	474	171.79	58.30	.32
Proportion Correct	473	.91	.08	474	.91	.08	.03
Number Memory							
Final Response Time Mean	473	158.73	43.01	472	147.48	39.98	.26
Input Response Time Mean	473	144.58	67.13	472	118.73	37.21	.39
Operations Pooled Mean	473	236.68	91.35	472	195.73	75.16	.44
Proportion Correct	473	.90	.09	472	.90	.09	-.03
Movement Time Pooled Mean[3]	479	32.16	8.15	487	33.21	8.05	-.12

[1] $\text{Gain Score} = \dfrac{\text{Time 2 Mean} - \text{Time 1 Mean}}{\text{Time 1 Standard Deviation}}$

[2] Gain Scores have been appropriately reflected so that positive gain scores signify improvement and negative gain scores signify decrement in performance.

[3] Movement Time is pooled across Simple Reaction Time, Choice Reaction Time, Short Term Memory, Perceptual Speed and Accuracy, and Target Identification.

TOW data: Predicting Event 5* with 14 Project A Predictors, ASVAB, and Combinations Thereof

Source: Jay Silva, ARI (1989, unpublished)

Predictor	r with Event 5	p	n	p unique resid var	stepwise R2	p/order
ASVAB						
AFQT (uncorrected)	.145	.0063	353			
CO composite (uncor) (CS+AR+AS+MC)	.180	.0007	353			
All subtests (optimal combo, uncorrected)**	.203	.1029	353			
Project A						
14 Project A tests below (optimal combo w'out removing ASVAB related variance)	.407	.0001	314			
14 Project A tests (optimal combo to predict ASVAB residualized TOW)	.354	.0002	311			
Computerized tests						
Cannon Shoot (Mean abs time err)	.159	.0029	348			
Short Term Memory (Mean trimmed decision time)	.050	.3533	350	.024	.0095	.0757/2
Number Memory (Mean pooled response time)	.069	.1954	349	.894		
Perceptual Speed and Accuracy (Mean trimmed decision time)	.052	.0392	352	.626		
Target Identification (Mean trimmed decision time)	.174	.0012	344	.091	.0125	.0403/3
Target Shoot (Mean log[dist+1])	.112	.0386	340	.824		
One-hand Tracking (Mean log[dist+1])	.206	.0001	352	.842		
Two-hand Tracking (Mean log[dist+1])	.294	.0001	353	.003	.0725	.0001/1
Spatial (# correct)						
Assembling Objects	.082	.1305	338	.103	.0088	.0844/4
Map	.161	.0030	338	.694		
Mazes	.108	.0482	338	.200		
Object Rotation	.168	.0018	344	.281		
Orientation	.242	.0001	340	.041	.0117	.0453/5
Reasoning	.151	.0050	344	.562		
ASVAB combo plus Project A combo	.421	.0001	311			

* Based on data gathered in 1987. From 1987 to 1988, the first qualifying table was changed from Event 5 to Event 3.

** MC was the only subtest accounting for significant ($p < .05$) unique variance. All others had $p > .4$.

TOW data: Predicting Event 3* with Four Project A Predictors, ASVAB, and Combinations Thereof

Source: Jay Silva, ARI (1989, unpublished)

Predictor	r with Event 3	p	n	SEM
AFQT (uncorrected)	.146	.0001	1,747	
9 ASVAB subtests (unit wtd combo, no corrections)	.191	.0001	1,747	
One-hand Tracking	.290	.0001	245	
Two-hand Tracking	.332	.0001	245	
Mazes	.020	.754	245	
Orientation	.162	.0113	245	
9 ASVAB subtests (optimal combo, no corrections)	.266	.0001	1,747	
Four Project A tests (optimal combo without regard to ASVAB)	.374	.0001	245	

Contribution to model:	p for unique variance
One-hand Tracking	.54
Two-hand Tracking	.005
Mazes	.041
Orientation	.035

Predictor	r with Event 3	p	n	SEM
Four Project A tests (optimal combo for predicting residual after ASVAB-related variance is removed)	.308	.0001	245	

Contribution to model:	p for unique variance
One-hand Tracking	.43
Two-hand Tracking	.029
Mazes	.010
Orientation	.817

Predictor	r with Event 3	p	n	SEM
ASVAB combo plus Project A combo	.421	.0001	245	

Note:
Best model for predicting ASVAB residualized variance, based on highest adjusted R^2, best c(p), and forward entry stepwise regression: Two-hand Tracking + Mazes.

*Based on data gathered in 1988. From 1987 to 1988, the first qualifying table was moved from Event 5 to Event 3.

UCOFT data: Predicting the Speed/accuracy Composite With Four Project A predictors, ASVAB, and Combinations Thereof

Source: Jay Silva, ARI (1989, unpublished)

Predictor	r with Speed/ Accuracy	p	n
AFQT (uncorrected)	.284	.0001	498
9 ASVAB subtests (unit wtd combo, no corrections)	.323	.0001	498
One-hand Tracking	.484	.0001	394
Two-hand Tracking	.504	.0001	394
Mazes	.308	.0001	326
Orientation	.328	.0001	326
9 ASVAB subtests (optimal combo, no corrections)	.480	.0001	498
Four Project A tests (optimal combo without regard to ASVAB)	.501	.0001	260

Contribution to model:	p for unique variance
One-hand Tracking	.41
Two-hand Tracking	.004
Mazes	.015
Orientation	.024

Predictor	r with Speed/ Accuracy	p	n
Four Project A tests (optimal combo for predicting residual after ASVAB-related variance is removed)	.312	.0001	259

Contribution to model:	p for unique variance
One-hand Tracking	.28
Two-hand Tracking	.121
Mazes	.025
Orientation	.17

Predictor	r with Speed/ Accuracy	p	n
ASVAB combo plus Project A combo	.565	.0001	259

Note:
Best model for predicting ASVAB residualized variance, based on highest adjusted R^2, best c(p), and forward entry stepwise regression includes at least Two-hand Tracking and Mazes. There is mixed support also for Orientation.

Table 6.3

<u>Correlations Between Computer Test Scores and Previous Experience With Video Games</u> (N = 250)[a]

Computer Test	Test Score	Correlation[b]
Simple Reaction Time	Mean RT	.12*
Choice Reaction Time	Mean RT	.15*
Perceptual Speed & Accuracy	Percent Correct	-.01
	Mean RT	.01
	Slope	-.03
	Intercept	.06
Target Identification	Percent Correct	.08
	Mean RT	.05
Short-Term Memory	Percent Correct	.13*
	Mean RT	.08
	Slope	-.16*
	Intercept	.18*
Number Memory	Percent Correct	.08
	Mean RT	.00
Cannon Shoot	Time Error	.18*
Target Tracking 1	Mean Log Distance	.22*
Target Shoot	Mean Time to Fire	.10
	Mean Log Distance	.27*
Target Tracking 2	Mean Log Distance	.16*

[a] Varies slightly by test.

[b] Correlations of .12 or greater are statistically significant at the .05 level, two-tailed test of significance. Signs of correlations have been reflected, where appropriate, so that greater video experience shows positive correlation with better test performance.

Correlations with Video Game-Playing Experience

Table 6.3 shows correlations of the 19 computer-administered test scores with the subject's previous experience playing video games. In the computer-administered tests, the question was asked: "In the last couple years, how much have you played video games on arcade machines, home video games or home computers?" Subjects selected one of the following five answers: "You have NEVER played video games," "You have tried a few games, but have generally played less than once a month," "You have played several times a month," "You have played at least once or twice a week," "You have played video games almost every day." These answers were given numeric values from 1 to 5, respectively. The mean score on this question was 2.99, SD = 1.03 (N = 256) and the test-retest reliability was .71 (N = 113).

Source: Peterson, 1987

Concurrent Validity Data Analysis

Time Used to Read Instructions and Complete Test Items For Computer-Administered Tests

Test	Period	N	Mean	Deviation
Demographics:	Instruction	9273	0 mins 19 secs	0 mins 40 secs
	Test Items	9273	4 mins 1 sec	1 min 13 secs
	Total		4 mins 20 secs	1 min 33 secs
Simple Reaction Time:	Instruction	N/A	N/A	N/A
	Test Items	N/A	N/A	N/A
	Total	9271	2 mins 20 secs	0 mins 58 secs
Choice Reaction Time:	Instruction	9272	1 min 0 secs	0 mins 22 secs
	Test Items	9272	1 min 58 secs	0 mins 22 secs
	Total		2 mins 58 secs	0 mins 44 secs
Short Term Memory:	Instruction	9269	2 mins 41 secs	0 mins 47 secs
	Test Items	9269	4 mins 54 secs	0 mins 40 secs
	Total		7 mins 35 secs	1 min 27 secs
Target Tracking 1:	Instruction	9256	3 mins 37 secs	0 mins 32 secs
	Test Items	9257	3 mins 47 secs	0 mins 2 secs
	Total		7 mins 24 secs	0 mins 54 secs
Perceptual Speed & Accuracy:	Instruction	9266	1 min 44 secs	0 mins 44 secs
	Test Items	9266	3 mins 25 secs	0 mins 34 secs
	Total		5 mins 9 secs	1 min 18 secs
Target Tracking 2:	Instruction	9254	2 mins 14 secs	0 mins 33 secs
	Test Items	9253	3 mins 47 secs	0 mins [illegible] sec
	Total		6 mins 1 sec	0 mins 34 secs
Number Memory:	Instruction	9256	3 mins 14 secs	1 min 11 secs
	Test Items	9256	6 mins 18 secs	1 min 15 secs
	Total		9 mins 32 secs	2 mins 26 secs
Cannon Shoot:	Instruction	9238	3 mins 8 secs	1 min 34 secs
	Test Items	9237	3 mins 43 secs	0 mins 5 secs
	Total		6 mins 51 secs	1 min 39 secs
Target Identification:	Instruction	9232	1 min 29 secs	0 mins 41 secs
	Test Items	9232	2 mins 34 secs	0 mins 34 secs
	Total		4 mins 3 secs	1 mins 35 secs
Target Shoot:	Instruction	9208	1 min 44 secs	0 mins 32 secs
	Test Items	9208	3 mins 6 secs	0 mins 39 secs
	Total		4 mins 50 secs	0 mins 59 secs
Total Time for Battery:	Total Test Time	9200	61 mins 3 secs	7 mins 14 secs

Note: Range for Total Test Time was 39 minutes 17 seconds to 2 hours 34 minutes 15 seconds.

Source: unpublished IPR materials by N. B. Peterson, et al., PDRI, 1986

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER — WP-RS-89-23

The Process and Results of Predictor Testing in TRADOC's Skills Selection and Sustainment (S3) Program

Clinton B. Walker
Army Research Institute

November 1989

REVIEWED BY:

Alan Drisko
Operations Research Analyst
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Encl 2

THE PROCESS AND RESULTS OF PREDICTOR TESTING IN TRADOC'S SKILLS SELECTION AND SUSTAINMENT (S3) PROGRAM

This report summarizes the events and outcomes of the testing component of the Training and Doctrine Command's (TRADOC's) Skills Selection and Sustainment (S3) Program. A short history of the program is followed by a description of the test package and support that the Army Research Institute (ARI) provided to the participating posts. Then, the testing operation is described post by post. Next, the outcomes of this program are evaluated in terms of process and impact. Following that, the potential benefits of administering the new tests in the training base and the pre-enlistment setting are compared. Finally, actual progress in moving the tests toward use for selecting/ classifying civilian applicants is described.

HISTORY

General Maxwell Thurman, then the Commanding General (CG) of TRADOC, initiated the S3 program for the purpose of improving operators' performance on weapons systems. His message on 11 December 1987 to the CGs of Forts Benning, Bliss, Knox, and Sill (Attachment 1) directed them to install new ARI tests of gunnery aptitudes and use scores on the tests to select trainees for gunnery training and positions. Among other things, he directed the testing of leadership potential and of spatial and psychomotor abilities for fast tracking and selection into gunnery specialties to start on 4 January 1988. GEN Thurman also directed that the CGs provide sustainment training for the active forces, but this paper does not deal with the sustainment training aspect of S3.

ARI was tasked to provide software, hardware, and technical assistance to get the testing going. The informal guidance to ARI was to create a test battery of no more than an hour that would include the Assessment of Background and Life Experiences (ABLE), a test of "will do" traits, along with the new tests of abilities.

At the kickoff meeting 15 Dec 87 at Fort Monroe, posts learned that TRADOC would not provide funds for the program. None of the participants, including ARI, was able to meet the 4 January suspense: Benning, Knox, and Bliss started testing mid- to late February, and Sill never started. ARI and the test sites reported weekly to TRADOC by telephone for most of 1988. Based on the results of the the pilot phase at Fort Bliss, which ended in July 1988, testing was not resumed. Early in 1989, MG Spigelmire, CG, Fort Benning, notified TRADOC that Benning was suspending testing for lack of resources. Benning resumed the testing in the summer at TRADOC's direction, even though funding was not provided.

ARI's TESTING PACKAGE AND SUPPORT

ARI provided the following materials and services to the field to support the testing: over 100 computer-based test stations, software for administering tests of one- and two-hand tracking, booklets for three pencil and paper tests (Mazes, Orientation, and ABLE), on-site training and computer tuneup just before the start of testing, a starter set of machine readable answer sheets, ongoing technical assistance (mostly by telephone), and a detailed Test Administrator's Manual. The computers were Seequa Chameleons and XLs that had had an average of 300 hours of use previously in the Army's Project A. By the start of S3, Seequa computers were no longer manufactured and the company with expertise to maintain them was no longer in business.

The Test Administrator's Manual was supplemented by a Test Users' Decision Guide for making personnel decisions with the test scores. In the guide, separate tables of norms were provided for ABLE and for a composite score derived from the spatial and psychomotor tests. These norms were based on large, comparable samples of examinees in Project A. In the composite score that was formed from separate scores on two spatial and two psychomotor tests, the psychomotor component was given a weight double that of spatial, because the psychomotor tests have a stronger relation with accuracy in gunnery.

Two decision rules were suggested in the Decision Guide, both intended to identify a larger pool of selectees than were needed to fill the relevant training slots. Two considerations led ARI researchers to build room for local discretion into these decision rules. First, we judged that any decision rule that removed local discretion completely would arouse resistance to the S3 program. Second, we believed that the users in the field have unique information on supply of and demand for gunners and on trainees' performance during Basic Training, both of which have a legitimate place in influencing selection of gunners.

One of the proposed decision rules weighted gunnery aptitudes (i.e., scores on the spatial and psychomotor tests) and leadership/motivation (scores on ABLE) equally; the other weighted gunnery much more heavily. These decision rules were offered to the users as examples, to be adjusted in light of local realities, and not as the last or exclusive word on selecting gunners.

Sites returned filled-in answer sheets and diskettes to ARI, where the data are now archived, and ARI recycled fresh diskettes to the field. The time for administering the five tests turned out to be about 75 minutes. For hand scoring the three pencil-and-paper tests, it took two persons about two hours to do 50 examinees. Late in July 1988, ARI converted ABLE from pencil-

and-paper to computerized form, which cut the labor of hand scoring by more than one half. In all, it took major commitments of time by 10 ARI Headquarters and Field Unit researchers plus contractor support to prepare and install the testing package.

THE PROCESS AT THE TRAINING POSTS

At the mutual request of ARI and the participating posts, TRADOC agreed for the testing to undergo a shakedown and evaluation period before the scores would be used for making personnel decisions. The three posts that conducted the testing made use of that grace period.

At the start of each testing session, examinees were told the following: the test scores might be used to pick a specialty or type of training for them; the scores would go into their personnel records; and, therefore, it was in their best interests to do the best that they could on the tests. In other words, the examinees were told that these tests counted.

Fort Benning

The staff of the Reception Battalion administered the tests, scored them, transmitted the scores to Headquarters, U. A. Army Infantry Training Center (ITC), and mailed the completed answer sheets and filled diskettes back to ARI. Insofar as possible, all Regular Army 11X receptees were tested. Toward the end of the AIT phase of training, the 29th Infantry Regiment conducted qualification testing on all 11H students. That testing covered, among other things, gunnery performance on a high fidelity TOW simulator at a gunnery range.

The results of S3 testing were continuously monitored by the Analysis and Studies Office of the Directorate of Training and Doctrine (DOTD) in the Infantry School. Researchers in that office gathered the spatial/psychomotor test scores from the Reception Battalion and, up to spring 1989, drove out to the gunnery range periodically to gather the data from the gunnery exercises. Their analyses and briefing in June 1988 led the Commander of the Infantry Training Center to approve use of the spatial and psychomotor test scores for selecting 11X Basic Trainees into 11H. Although it was not approved for use in making personnel decisions, ABLE continued to be administered in the S3 package. The Analysis and Studies Office also designed forms for recording 11H trainees' scores on selected hands-on tasks (e.g., prepare range card; assemble weapon), worked with the 29th Infantry Regiment to set up a procedure for filling out the forms, gathered the completed forms, and mailed them to ARI. By the end of the S3 Program, the new tests had been taken by about 18,000 receptees at Fort Benning. The ARI Field Unit at Fort Benning supported the S3 program throughout by keeping the participants at ARI Headquarters and Fort Benning informed of perceptions and realities at both ends of the operation.

Fort Knox

At Knox, the new tests were housed at the Reception Battalion and run almost to the program's end by the SQT Office of the Armor School. Scores on the S3 tests were transmitted to the Armor Training Brigade, the Office of the Chief of Armor, the ARI Field Unit on posts, and, by mail, to ARI Headquarters. Examinees' gunnery scores on the Institutional Conduct of Fire Trainer (ICOFT) later in training were gathered and evaluated by the local ARI Field Unit. Their analyses led the Fort Knox CG, MG Tait, to agree in September 1988 to permit scores on the spatial and psychomotor tests to be considered, as one of several types of evidence, in selecting volunteers into the Excellence in Armor Program. ABLE continued to be administered along with the ability tests. By the end of the S3 Program, about 7,500 receptees had taken the new tests there.

Fort Bliss

Here, the ARI Field Unit conducted the predictor testing on post and the performance testing at White Sands Missile Range. At the direction of the Air Defense Artillery School, the test package included only the psychomotor and spatial tests. The examinees were students in 16S and 16P who were close to finishing the AIT phase of their OSUT. On one day they took the S3 tests, and within the next two days were tested on high fidelity simulators of their respective weapons systems. A team of researchers at ARI's local Field Unit and Headquarters evaluated and reported the data. Based on an absence of positive results at Fort Bliss, the testing was not continued beyond the pilot phase. Only 101 cases were tested there owing to the very limited funding for gathering performance data on the Realistic Air Defense Engagement System (RADES).

Fort Sill

ARI held meetings with the Field Artillery Center and School at Fort Sill in August and November 1988, and again in February 1989 when the test package was installed. Late in 1988, the Field Artillery Center and School decided to include four MOS in its S3 testing, each with a unique set of spatial and psychomotor tests. ARI designed an expanded test package and installed it at Fort Sill's Reception Battalion in February 1989. It included five pencil and paper tests of spatial abilities and software to administer as many as seven computerized tests, the exact choice of tests varying with MOS. By the close of the S3 Program, repairs had not been completed on the building that would have housed the testing operation, and so testing never got under way.

DESCRIPTION AND EVALUATION OF THE OUTCOMES

Test Administration

The team of field and ARI personnel succeeded in setting up, starting, and maintaining a successful testing operation for eighteen months. Over 25,000 entry-level students took the new S3 tests. Given that the program had been launched suddenly, at the participants' expense, and with aging hardware, the posts' effective conduct of the testing program was a considerable achievement. The effectiveness of the testing is proven both by the numbers tested and by the findings that the new tests work well for identifying students with aptitudes to be good tank and anti-tank gunners, as discussed next.

Test Validities

In the S3 Program, the high validities of spatial and psychomotor tests which had been found in 1986-7 with ARI doing the testing, were replicated; the tests' power to predict levels of performance on simulators of tank and anti-tank gunnery held up in the S3 setting (Smith & Walker, 1988; Graham, 1989). In MOS 11H, students who scored in the upper one-third in S3 gunnery aptitudes (i.e., on the norms of spatial and psychomotor tests) attained Expert or First Class levels of accuracy on the TOW simulator at a rate 12% higher than the prevailing average. Seventeen percent more of those in the top third attained the top two levels of gunnery qualification than those in the bottom two-thirds. A small (N = 59 students) live fire exercise at Fort Benning in June 1988 confirmed those findings, with 11% more of the upper third (in spatial/psychomotor abilities) hitting the moving target 2,000 yards down range on their single shot than those in the lower two-thirds. Although that result was not statistically significant, in the context of the other TOW data, it is entirely credible. Students in the upper third of S3 gunnery aptitudes also trained up to the qualifying level faster: 41% of them qualified on the earliest possible gunnery table, while only 12% of the others did. The exact figures varied a bit as time passed, but the advantage of the high spatial/psychomotor students endured.

An additional benefit was gained from the performance data that the Studies and Analysis Office at Fort Benning collected at the firing range: ARI's analyses of these data for TRADOC's DCSRM have been included in the report on the Soldier Performance Research Project (Block, Polich, & Horne, in press) to confirm that quality, as defined by scores on ASVAB, has a positive impact on both firing accuracy and hands-on tasks on the TOW.

The results of S3 at Fort Knox that are reported here are from the work of Dr. Scott Graham of the ARI Field Unit there

(Graham, 1988, 1989). In 19K, students who scored in the upper third in S3 Gunnery aptitudes had a hit rate 7% higher on the ICOFT than the prevailing norm, and 13% higher than the lower two-thirds. That finding exactly replicates the differences in a sample of officers in earlier research (Smith & Graham, 1987). To equal the hits of these high ability students, 11% more tanks manned by average gunners would be needed.

In comparison with students in the Excellence in Armor (EIA) Program, who averaged 71% hits on the 20th hour of ICOFT, all 19Ks in the upper third of S3 gunnery aptitudes averaged 70%. Even on the input side, the EIA selection system is capturing some of the same information as the new tests: EIAs scored significantly higher on the spatial and psychomotor tests than the non-EIAs (Graham, 1989, p. 14). However, there is still a strong correlation of S3 test scores with the ICOFT scores of the EIA students. That finding means that scores on tests of spatial and psychomotor abilities provide sufficient unique information to aid in selecting EIAs so as to improve their gunnery performance.

On ABLE as well, EIAs score significantly higher than non-EIAs who were tested in S3. The non-EIAs' average ABLE score fell at the 54th percentile in the norms, while the EIAs' fell at the 68th percentile. Thus, the EIA selection process operates as though it were selecting from the upper one-third on ABLE.

At Fort Bliss, scores on the new tests were not related to trainees' performance in gunnery on the Realistic Air Defense Engagement System (RADES). ARI and the Air Defense Artillery School agreed that the two air defense weapons systems in the S3 testing, Stinger and Chaparral, probably do not require the abilities measured by new ARI tests nearly as much as they require other abilities (e.g., vision).

Credibility of the Results

Based on data from the first few months of testing, leadership at Benning and Knox agreed to permit the test scores to enter into the decision process for selecting the relevant categories of students. Thus the tests and the evidence of their effectiveness were credible to users in the field. The negative results at Fort Bliss were credible as well, both to the field and to ARI. So the test results provided good information for evaluating the utility of the test package at all three posts.

Because the S3 tests have repeatedly shown such high potential benefit to the Army for tank and anti-tank gunnery, ARI has briefed the results of the program to a variety of senior decision makers in the Army and DoD. These include the following General Officers and Senior Executive Service officials: CG 8ID (MG Waller, Jul 88), CG V Corps (LTG Woodmansee, Jul 88), CG 3AD (MG Joulwan, Jul 88), CG Fort Knox (MG Tait, Sep 88), DCSPER of

the Army (LTG Ono, Sep 88), VCSA (GEN Brown, Sep 88), DoD's Director of Military Accession Policy (Dr. Sellman, Dec 88), the Assistant (to DoD's Director of Environmental and Life Sciences) for Training and Personnel Systems Technology (Dr. Alluisi, 12/88 and 6/89), CG Fort Benning (MG Spigelmire, Feb 89), DoD's Assistant Military Deputy for Research, Development, and Acquisition (MG Beltson, Jul 89), ODCSPER's Deputy for Military Personnel Management (BG Stroup, Aug 89), and the ADCSPER of the Army (MG Budge, Sep 89).

Impact on Selecting Gunners

One major aspect of the S3 testing operation was not a success: the test scores did not have an impact on actual personnel selection. Four components of a system for effective use of the scores were not uniformly installed. First, at Fort Knox, the post-level information management systems for getting the test scores to decision makers were not set up and maintained. Test scores did get out of the Reception Battalions to the Training Brigade at Knox, but not out to the students' units. As a result, decision makers on-site were not able to use the test scores for selecting trainees based on aptitudes for gunnery.

Second, a system was not set up or maintained for getting the test scores or some sign of them, like an Additional Skill Indicator, into students' permanent personnel records. Thus, the students' future gaining units had no way of identifying individuals with high spatial and psychomotor abilities so as to make best use of their talents. In the case of the EIA Program, that missing link of information is less of a loss because fast tracking moves the EIAs ahead in the progression toward gunners' seats. But for 11Hs, the information on gunners' aptitudes now exists in an easy-to-retrieve form only in ARI's databases.

Third, even where the test scores had gotten into the hands of those who select gunners (viz., at Fort Benning), procedures for using the information effectively to make the relevant personnel decisions were not installed. Just getting the scores into the users' hands does not guarantee that the scores are usable or used; practical, believable procedures need to be available for decision makers to employ in consuming the test scores. At Fort Benning, even though the test scores did get to students' units, other factors drove assignment of 11Xs to specific Infantry MOS.

Fourth and equally important, the marketing, training, and monitoring necessary to institutionalize use of the test scores did not take place. Of these four components, probably only the first could have been effected in a relatively short time; but even that component would have required additional resources that were not available (e.g., networking hardware and software for the computers; machines for scanning the answer sheets at the

test sites). In summary, S3 was successful in generating useful scores on new tests, but it did not follow through with the institutional changes needed for getting the test scores used. Bringing about the changes would have required continuing active involvement together by TRADOC, ARI, and each post.

POTENTIAL UTILITY OF THE S3 TESTS IN THE TRAINING BASE

Now that a strong relationship between spatial/psychomotor abilities and gunnery performance has been replicated in the S3 setting, the question remains, "How could the benefit of the tests be realized?" In this section, the potential benefit of administering the tests at Forts Knox and Benning for selecting gunners is discussed.

Fort Knox

Even though it is designed as a leadership, not a gunnery, program, Excellence In Armor (EIA) appears to be doing an effective job of selecting trainees who have gunners' aptitudes without using scores on spatial or psychomotor tests. Still, the finding that EIA students' scores on the S3 psychomotor and spatial battery correlate significantly with their performance on ICOFT shows that the new tests could aid in selection into EIA. Whether the EIA program can continue to be as successful in producing superior gunners as the quality of accessions varies over time is unknown.

EIA selects students based on a broad set of indicators (ASVAB scores, academic performance, fitness, leadership, etc.) that accumulate by the end of the first seven weeks of training. It is not surprising that this "seven week test" provides good information on students' leadership potential. Also, the "seven week test" has the advantages of making use of ASVAB scores and direct observation (i.e., two persuasive types of evidence), it is a local product, it is in place, and it appears to enjoy strong support on post.

The choice between using ABLE scores and the "seven week test" is a false choice: ABLE scores could be added into the set of evidence that is examined in selecting volunteers for fast tracking. When ABLE scores from pre-enlistment testing (see below) become available in receptees' personnel records, then they could be used as easily as ASVAB scores are now in selecting volunteers into EIA.

Although there is considerable evidence that the S3 tests have good information for use in selecting EIAs, the question remains as to the <u>added</u> utility of scores from the S3 tests. It is not obvious that the S3 tests add enough new information to warrant the cost of the testing program at Fort Knox. The EIA program is different from settings where gunners enter their duty

positions essentially unselected for special gunnery aptitudes.

One relevant research question for TRADOC and ARI is how widely and systematically scores on Conduct of Fire Trainers are used by TO&E units to select and de-select gunners. If units use COFT scores for putting and keeping individuals in gunners' seats, then they are selecting on the basis of special gunners' aptitudes.

Fort Benning

At Fort Benning, the present system for classifying 11Xs into the four Infantry MOS does not have the effect of taking spatial and psychomotor abilities into account. As a result, students with the full range of spatial and psychomotor abilities enter 11H, and their average performance on the M-70 TOW simulator is markedly lower than it could be. Using the new ability tests as the primary selector of 11Xs into 11H would significantly improve TOW gunnery performance.

But even if the administrative ground work and support for using the test scores were in place, three aspects of personnel processing at Benning would have to be dealt with before the new ability tests could have a significant effect on assignment to MOS. First, the distribution of high quality students (in terms of the ASVAB General Technical score) across MOS is constrained by a locally developed system for spreading quality across MOS. Second, some students enter OSUT at Fort Benning with enlistment contracts that make it necessary to put them in MOS 11H. There is no paperwork mechanism now that allows post-enlistment tests to change those contracts for students with low spatial and psychomotor abilities. Finally, occasional surges in demand for TOW gunners make it necessary to channel students into 11H irrespective of their aptitudes for gunnery.

There may be ways to incorporate scores on spatial and psychomotor tests into the local system for distributing quality soldiers and the correlation of ASVAB scores with the gunner's aptitudes is low enough that 11H could be filled without 11H having to monopolize the high-ASVAB students. But the other two constraints, enlistment contracts and surges in demand, seem inescapable. So the possible impact of the new tests at Fort Benning is limited to some unknown degree by the realities of personnel processing for 11Hs.

MAXIMIZING THE UTILITY OF THE S3 TESTS

A solution to these limits on the impact of the new tests at Forts Benning and Knox has occurred to personnel at both the training schools and ARI: install the tests in pre-enlistment processing for selecting and classifying civilian applicants. By siting the tests in the pre-enlistment setting, the Army would be able to identify persons with gunners' aptitudes from the pool of

all applicants, and would be able to assign qualified individuals to all of the MOS that require spatial and psychomotor abilities. Since about 400,000 Army applicants take ASVAB annually, and accessions average about 115,000 annually, the number of high spatial and psychomotor individuals who could be assigned to gunnery specialties would be vastly larger than the numbers now entering 11X and 19K. Perhaps just as important, the numbers of persons with low aptitudes for gunnery who could be steered away from gunnery occupations would be enormous. Such a system could raise the level of gunners' aptitudes in the 11Hs and 19Ks by doing the person-job match before training, which was the original intent of S3.

The benefits to the Army would be improved trainability and performance in critical skills for relevant MOS. These benefits to TRADOC and the active forces would be attained, however, at some cost to MEPCOM. For this reason, implementation of such new ability tests at MEPCOM sites will not occur until the cost-effectiveness of installing a computerized testing system has been evaluated and found positive.

At this writing, the Department of Defense is preparing such an evaluation. In 1990-91, the DoD Enhanced Computer Assisted Testing (ECAT) Project will be conducted to evaluate a battery of Army and Navy computerized tests, including the four S3 ability tests. For that project, several pencil-and-paper tests of spatial abilities will be converted to the computerized medium. Data collection will be carried out in the training bases of all of the Services, but hardware, software, and staffing will all be supplied to sites via an existing Navy contract. Upon receipt of details from DoD, ODCSPER will open formal coordination with TRADOC on this new project.

Overall, the objective of ECAT is to see whether it would be cost-effective to computerize all pre-enlistment testing. The new tests are to be evaluated as supplements to ASVAB, not as replacements. As a final note of evaluation on S3, results from the S3 program have influenced DoD's choice of tests for ECAT; so although the S3 testing is ended, its impact continues in ECAT's pre-enlistment tests of the future. The potential long run benefit of those tests - across all relevant specialties in all Services - is very great.

As for testing the will-do attributes which ABLE measures, administering ABLE in the pre-enlistment setting would have the same advantages of reaching the full applicant pool and helping assignment to all MOS. DoD has included a form of ABLE in its Applicant Screening Profile (ASP), which has already been printed by MEPCOM. At present, ASP awaits final approval for use in testing applicants to all of the Services, and so is on the verge of implementation.

The ASP is intended primarily as a selection screen to reduce attrition, but economists in ARI's Manpower and Personnel

Policy Research Group have developed strategies for using ABLE for classification (i.e., person-job match) as well. Once the amount of attrition is reduced by screening out the highest risk applicants, the cost of attrition can be reduced further by distributing the highest risk accessions across MOS where the cost of training is not high.

REFERENCES

Block, Polich, & Horne (in press). Soldier Performance Research Project. U.S.A. TRADOC DCSRM Report #___.

Gast, I. F. & Johnson, D. M. (1988, December). Evaluating psychomotor and spatial tests for selecting Air Defense Gunners. Paper presented at the annual meetings of the Military Testing Association, Arlington, VA.

Graham, S. E. (1988, December). Selecting soldiers for the Excellence in Armor Program. Paper presented at the annual meetings of the Military Testing Association, Arlington, VA.

Graham, S. E. (1989). Assessing the impact of psychomotor and leadership selection tests on the Excellence in Armor Program. (ARI Research Report 1510). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Smith, E. P. & Graham, S. E. (1987). Validation of psychomotor and perceptual predictors of Armor Officer M-1 gunnery performance. (ARI Technical Report No. 766). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Smith, E. P. & Walker, M. R. (1988, December). Testing psychomotor and spatial abilities to improve selection of TOW gunners. Paper presented at the annual meetings of the Military Testing Association, Arlington, VA.

SELECTION AND CLASSIFICATION TECHNICAL AREA

WORKING PAPER WP-RS-89-16

Concurrent Validation Codebook - 64C

Winnie Y. Young
Kathy Cordray Austin
Jeffrey J. McHenry
Lauress L. Wise
American Institutes for Research

Henry H. Busciglio and Elizabeth J. Brady
Selection and Classification Technical Area

August 1989

REVIEWED BY:

Jacinto M. Silva
Research Psychologist
Selection and Classification
Technical Area

APPROVED BY:

Michael G. Rumsey
Chief
Selection and Classification
Technical Area

CLEARED BY:

for Curtis L. Gilroy, Acting Director

MANPOWER AND PERSONNEL RESEARCH LABORATORY

U. S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, Virginia 22333-5600

Introduction

This codebook is being provided to the Defense Training and Performance Data Center (TPDC) in conjunction with the transmission of data tapes to TPDC. These tapes contain data from the Army's Project A, a large-scale multiyear research effort designed to improve the selection, classification and utilization of Army Enlisted Personnel. The initial large data collection effort of Project A was the Concurrent Validation (CV) in which predictor and performance data were collected concurrently from soldiers after 12 to 27 months of service. The CV resulted in the collection of more than fifteen million data points from approximately 9000 soldiers in 19 Military Occupational Specialties (MOS).

In order to effectively use the large amount of data collected in the field during the CV, as well as variables obtained from other sources, a Longitudinal Research Data Base (LRDB) was established. Due to the requirements of various types of data analyses to be performed, it was critical that the files in the LRDB be efficiently designed and well documented. Generally, item and step level data are maintained on tapes and merged with macro level variable files as required. The large separate MOS files for the CV contain the primary macro level variables used for analysis including Task Scores, Scale Scores, Construct Scores, Mean Ratings Scores, Imputation Flags and Residual Scores as well as demographic, accession and Army-collected variables.

This paper presents one of the Concurrent Validation Codebooks developed by the LRDB manager and her team for the nine MOS for which Job Knowledge Tests, Hands-On Tests and MOS-specific Ratings were developed.